To what extent are language assessment scores used in valid and ethical ways for the purpose of immigration, employment and professional registration in a New Zealand context?

By Linda Edwards

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Canterbury

College of Education, Health and Human Development

2019

ABSTRACT

High-stakes language tests are used by immigration authorities, professional registration bodies and employers to make decisions about the employability of non-native speakers of English in English speaking countries. Additionally, native speakers of English are required to take high-stakes language tests to gain immigration points or for professional registration purposes. However, there is a paucity of research on the perceptions of native speaker test takers, non-native speaker test takers and employers regarding the test content and constructs being measured in English language tests compared with the language used in the contemporary workplace. This thesis, therefore, addresses the research question: To what extent are language assessment test scores used in valid and ethical ways for the purpose of immigration, employment and professional registration in a New Zealand context? There has also been little research into language test validity using an argument-based approach. This research study attempts to fill this void by applying an argument-based validation approach to a high-stakes language test, the International English Language Testing System, which is currently widely used for employment, professional registration and immigration purposes in New Zealand. Four sub questions were developed as the research progressed to encompass domain, extrapolation, decision, and consequence inferences. The data which provided evidence for these inferences were collected from semi-structured interviews with stakeholders who had taken or used the test for employment purposes. Their perceptions provided support or rebuttal for the inferences in the validation argument. The ensuing results helped build an argument which clarifies the extent to which language test results are being used in valid and ethical ways for employment purposes.

The perceptions of the non-native speaker test takers, and particularly the native speaker test takers and employers highlight the social, economic and political implications of using language tests to infer the communicative language competence of migrants in three employment sectors: education, business/IT and manufacturing/retail. Recommendations are made for the future use of language tests which assess higher language cognition and communication skills as opposed to basic communicative language competence. The insights gained from this research demonstrate that native speakers of English support the testing of oral language skills, but that the ways in which written language skills are used in the contemporary workplace have changed. As a result, there is a strong need for language test developers and test users to engage in collaborative discussions in order to ascertain which types of current and future written language skills should be tested for migration and employment in English-speaking countries.

ACKNOWLEDGEMENTS

For Akiko

To the William and Ina Cartwright Scholarship Trust. Their financial assistance was invaluable.

To Una Cunningham, John Boereboom and Missey Morton for inspiring me. This journey would not have been possible without their initial encouragement, stimulation and belief in the importance of investigating the validity and ethics of language tests. Thank you for challenging me and providing essential critical feedback, which was invariably thought-provoking.

To Jo Fletcher, Alison Arrow and Jocelyn Howard for keeping me on track. Many thanks especially to Jo Fletcher for the time and effort she put into providing feedback.

To all of Una Cunningham's LATL labbers. I enjoyed and appreciated our lively debates and shared presentations about language acquisition, language teaching, and the use of technology. Your own research experiences and confidence in me helped me immensely.

To my husband, Rob, for his constant love, support, reassurance and advice. You are my rock.

TABLE OF CONTENTS

Academic Module	The academic module of the International English Language Testing System (IELTS) test. This is used for entrance to higher education institutions, for professional registration purposes and by some employers.
BLC	Basic language cognition. As defined by Hulstijn (2011; 2015), this is the language used by the majority of native speakers of a language who do not have high levels of educational or professional experience.
CEFR	The Common European Framework of References for language proficiency.
EAP/ESP	English for academic purposes/ English for specific purposes.
General Training	The general training module of the International English Language Testing System (IELTS) test. This is used for entrance to some vocational courses, for immigration and employment purposes.
HLC	Higher language cognition. The language used by native speakers of a language who have had prolonged experience of using language in academic and professional spheres.
iBT TOEFL	Internet-based version of the Test of English as a Foreign Language. This is used for university entrance purposes.
IELTS	The International English Language Testing System.
INZ	Immigration New Zealand.

IUA	Inference Use Argument framework for test validation. This framework was created by Michael T. Kane from the Educational Testing System in the United States.
LAL	Language assessment literacy.
NSs	Native speakers of a language. Speakers for whom the language is their mother tongue.
NNSs	Non-native speakers of a language. Those who have learned a language in addition to their mother tongue.
PTE	Pearson Test of English.
Test developers	Organisations who have created and developed a test.
Test takers	Candidates who take a test.
Test users	Employers and decision makers who use the results of test to make decisions about test takers.
TLU	Target language use domain. This is the context in which the language is to be used. For the purposes of this study, the target language use domain is the contemporary New Zealand workplace.

CHAPTER 1

INTRODUCTION

1.0 Introduction to the thesis

This thesis explores the meaningfulness and consequences of language testing for migrants to New Zealand who are required to take a high stakes language test such as IELTS (International English Language Testing System), TOEFL (Test of English as a Foreign Language), TOEIC (Test of English for International Communication), the Cambridge certificate exams or the newer PTE (Pearson Test of English). There are three main groups of stakeholders who are directly or indirectly affected by language testing and the subsequent test score results: test developers, test takers and test users (employers and policy makers). The core aim of this research was to provide insights into the perspectives of native speaker as well as non-native speaker test takers, and native speaker employers who employ a variety of nationalities in New Zealand, towards the language assessed in high-stakes language tests compared with language used in the workplace. The intention was to establish whether the content and constructs which are assessed in a high-stakes language test are convergent or divergent with language competence for contemporary workplaces.

While there has been research dedicated to the viewpoints of non-native speakers who have taken a language test for migration purposes (Gribble, Blackmore, Morrissey & Capic, 2016; Roshid & Chowdhury, 2013), the perspectives of native speakers of English (test takers and employers) provide important, yet under-researched, facets for debate in the field of language testing and helped build an argument for the validation of using test score results for the purposes of immigration, employment, and professional registration in New Zealand.

1.1 Motivation for this research

Prior to starting this study, I had learned European languages at university for my first degree and consolidated my understanding and use of these languages by living and working in many European countries. I was therefore aware that learning a foreign language at school does not relate to language use in context and that language competence can only be enhanced by extensive, authentic output.

Since migrating to New Zealand in 1990, I have worked as an English language teacher, Director of Studies, Director of a teacher training institute, IELTS examiner (since 1997) and on the MTESOL programme at a New Zealand university. In addition, I established English for Employment courses for professional migrants seeking employment in New Zealand. These

migrants were doctors, engineers, teachers, vets and other skilled migrants from areas as widespread as Africa, Asia, Russia and the Middle East. This has helped me understand language learning from different contexts, and the role of a person's mother tongue, identity and socio-educational background in determining successful trajectories towards social and professional inclusion in a new country.

In my experience, many skilled professionals resorted to working in cafes, shops and petrol stations as their language skills and previous qualifications were not recognized in New Zealand, a situation reported by other researchers (Gribble et al., 2016; Roshid & Chowdhury, 2013; Rumsey, Thiessen, Buchan & Daly, 2016). At the same time, as an IELTS examiner I have come across native speakers of English who have needed to take language assessment in English and not gained the score they wanted. This piqued my interest into gathering information on the perspectives of native speaker test takers as well as non-native speaker test takers who live and work in New Zealand. The perspectives of test users, in this case employers, who make decisions about the employability of migrants are important for the debate about language test validity (Kane, 2013; Kunnan, 2018), particularly since it seems to be a waste of human resources if migrants' skills and knowledge are not being effectively utilized.

1.2 New Zealand context and background

High-stakes language testing systems, such as IELTS, are being increasingly used in New Zealand as a part of the process of gaining employment and permanent residency (PR); there were two or three test centres in New Zealand in the 1990s but now there are around twenty. The issue, as noted by Kunnan (2018), is that assessments, and especially high-stakes assessments, are not challenged by test takers and local communities as there is often a perception that tests are developed by 'experts' and, therefore, cannot or should not be challenged. Further, stakeholders such as policy makers and employers who use these tests for making decisions about a person's suitability for education, work or immigration purposes are likely to accept the results of a language test if they perceive it to be reliable, but they may not know how to interpret the test results reliably (Caines, Bridglall & Chatterji, 2014).

Language assessment is a complex area. It is an aspect of acceptance to a country for work purposes that is often not thought about by many members of society, especially native speakers of English who have never needed to think about the acquisition of their mother tongue, what it means to 'know' a language (Ortega, 2013), the difference between basic language competence and cognitive

language proficiency (Cummins, 1984; Hulstijn 2011, 2015), nor indeed, their own ability to use their first language (L1) accurately. Yet, language assessment is used to allow entrance to the New Zealand workplace (in the case of professional bodies such as the Education Council) and to New Zealand society (INZ, 2018) while inferences are made about migrants' English competence according to scores that have been achieved in high stakes language tests (Kane, 2013). These uses of language assessment and inferences about a person who has English as an additional language (EAL) are generally made by monolinguals who may have little comprehension of second language acquisition or language assessment literacy (Merrifield, 2008; 2011; O'Loughlin, 2013; Pill & Harding, 2013). For the above reasons, the aim of this research was to investigate the inferences that are made about language competence and if language assessment is being ethically used in a New Zealand context.

1.3 The research problem

In New Zealand, as in many other countries, high-stakes language tests are now being widely used as gate-keeping devices for non-academic purposes. The test itself and the resultant scores are being used to serve as entry requirements in immigration policies and to ensure test takers have the necessary linguistic skills for the workplace. Both native speakers of English and non-native speakers of English are required to gain certain test scores to immigrate to and work in Australia, Canada, New Zealand and the United Kingdom. However, there has been very little research into the uses of test scores for the modern workplace and how much test performance relates to the type of communication needed in the workplace. Aryadoust (2013) recommended using the IELTS test for this type of research: "Given that this research enquiry has never been addressed by the IELTS developers, validating the test for non-academic purposes is timely and should be of high priority" (p.222). This research, therefore, aims to analyse the test takers' and test users' perceptions of the IELTS test and how it is used for employment purposes in New Zealand.

1.4 Migration to New Zealand

As regards entry requirements for migrants into New Zealand, Immigration New Zealand (INZ, 2018) lists eight major occupational groups on the long-term skill shortage lists: construction, engineering, finance/business, health and social services, ICT, hospitality (chefs), science and trades. The region of Canterbury has its own specific skills shortage due to the Christchurch post-earthquake rebuild with a distinct focus on construction, engineering, telecommunications and trades staff (INZ, 2018).

A change in immigration rules in 2016 has meant that the IELTS band score required to apply to Immigration New Zealand as a skilled migrant for permanent residence is 6.5 (it was previously 5.5) in either the Academic or General Training module. The list of occupations which are registered on the immediate skill shortage list (INZ, 2018) is long and there is considerable variation in the levels of education and training needed for each one (from bakers and bricklayers to cardiologists and university lecturers). Although these groups of people are required to obtain a band score of 6.5 across all four skills, this is the same band score that is needed for an international student to study a PhD in New Zealand. Setting the same test score prerequisites for such a wide range of jobs and academic backgrounds appears to be incongruous since construction workers, chefs, sheep shearers and accountants or lawyers all need an IELTS score of 6.5. This led me to question the communication skills that such a range of workers need on the job and if IELTS tasks are testing relevant aspects of communicative competence. Indeed, Merrifield (2008) researched the impacts of using IELTS as an entry criterion for professional associations in Australia, New Zealand and the USA and ultimately recommended that test validation be undertaken which probes the language skills needed in the workplaces of the professional associations which require IELTS test scores. However, Merrifield's (2008) study did not encompass language use in today's workplace.

The majority of the professional bodies in New Zealand require an IELTS score of 7.0 in every skill, as documented by Merrifield (2008). A recent issue in New Zealand has been the need to employ overseas teachers due to a general shortage of teachers. For this reason, teachers from different sectors were included in my research. The Teaching Council of Aotearoa New Zealand requires a score of 7.0 in the four skills, regardless of whether the teachers are in the early childhood or university sectors. Murray, Cross and Cruickshank (2014) noted the need to research the connections between language test scores and how effective teachers perceive themselves to be in the workplace.

Gribble et al. (2016), O'Loughlin (2013) and Pill and Harding (2013) found that employers and other test users in Australia had little knowledge of what IELTS assesses and what it does not assess. These Australian findings concurred with my own experiences as an English language educator. However, there appeared to be a dearth of research regarding the viewpoints of test users in the New Zealand context. It was, therefore, considered that the viewpoints of employers in a variety of occupational sectors in New Zealand would add to the picture of how test users use and interpret language test results.

The three sectors interviewed for this research study were education, business/IT and retail/manufacturing.

1.5 Reasons for using IELTS as an exemplar of a high-stakes language test

As stated in 1.1, there are several high stakes English language tests which are used as gatekeeping tools for study, work and migration purposes. It was beyond the scope of this study to investigate all of them, so for the purposes of efficiency the IELTS test is mostly referred to. This is because it was the most widely used test in New Zealand (at the time of writing) to ascertain immigrants' entry to New Zealand.

From the outset, I want to state that IELTS is used throughout as an exemplar of a high-stakes language test, which has the same advantages and disadvantages as the other major language tests, as described by Hulstijn (2011; 2015). They have all been developed by test developers who use credible international researchers to support the reliability of their testing systems. Yet, they focus on discrete item accuracy and complex language which may or may not be a reflection of language use in a wide range of English-speaking contexts. Further, these tests have evolved into commercial entities and are used as gatekeeping devices (Macqueen, Pill & Knoch, 2016) by educational, professional and political bodies. It has not been my intention to use IELTS as a scapegoat, an accusation made by Hall (2010) against Uysal's (2010) critical response to the use of IELTS by higher education institutes. I agree with Hall (2010) that Uysal (2010) misguidedly interpreted the IELTS acronym to mean the test developers claim the test is assessing international English, and that it is difficult, or impossible, for a test to take into account all the varieties of English. I concur, however, with Uysal (2010) that the IELTS writing tasks (as with other language tests) may focus overly on traditional lexico-grammatical accuracy with too much emphasis on mechanical features such as spelling and paragraphing, rather than meaning and intelligibility. I contend that both Hall (2010) and Uysal (2010) focused too much on reliability when mentioning issues related to the analytic scales and inter-rater correlation of the IELTS test. My intention has been to research test validity as opposed to reliability, and to heed the recommendation made by Hall (2010) by focusing on what he labels as the more difficult issues: the social, economic and political consequences of high-stakes language tests such as IELTS. This will, hopefully, help fill a gap in the research literature since Chalhoub-Deville (2016) and Lumley and Brown (2005) have reported that publications in this area are scant.

Using IELTS as an exemplar of a high-stakes language test has an added bonus because, as pointed out by Hall (2010), IELTS is an internationally popular test, which is rigorous in its design, delivery and marking procedures. In addition, the IELTS partners are proactive in resolving any issues with the validity of the test by continually calling for research and subsequently revising the test in line with such research. I believe, however, that Hall (2010) was incorrect to state that test users, such as universities, need to take more responsibility for understanding and interpreting test scores. While I agree that there does need to be a more critical debate over language standards (Hall, 2010), it may also be incumbent on test developers to engage in disseminating more language assessment literacy for test users.

1.5.1 Overview of the IELTS test

Tests such as IELTS have no pass or fail grades. The test scores are reported on a scale of 0.0 to 9.0. but the question is what do these numbers mean for test takers, test users and other stakeholders? According to Gribble et al. (2016), test users in Australia were unaware that IELTS candidates may receive half band scores and they found it difficult to interpret test scores.

Looking at the IELTS testing process, the exam comprises four modules which cover the four key language skills, often referred to as macro-skills: listening, reading, speaking and writing. There are two types of exam, the General Training (GT) and Academic (Ac) tests which are mainly used for migration and university entrance respectively. Nonetheless, the listening and speaking modules are the same for all candidates (for both General Training and Academic test takers). In fact, the listening section of the exam includes dialogues and monologues but with a focus in the last two sections on academic contexts, including a university lecture. The speaking section takes approximately 14 minutes and starts off with general personal questions (part one dialogue), followed by a two-minute speech (part two monologue) and subsequent questions of a more topical, abstract nature (part three dialogue).

The two sections of the exam which differ, depending on whether the candidate has taken the General Training or the Academic test, are the reading and writing parts. The content of the Academic reading test may originate from journals or magazines, but the topics are more suitable for candidates who want to enter university whilst the General Training test uses documents from an everyday environment with the recent addition of company documents. However, it is important to note that the types of questions (True/False/Not Given, no more than three words/ multiple matching) and, therefore, the constructs being tested are

identical. In the writing section, there are two tasks with the second one being similar for both the Academic and General Training modules, namely a discursive essay in response to a point of view, argument or problem. The main difference is that the response to the academic question should be more formal in register while Task One asks candidates to write a letter (General Training) or to describe a graph, table or diagram (Academic).

Despite the differences between the Academic and General Training modules, the rubrics that examiners follow to assess speaking and writing proficiency (Appendices O and P) are the same whichever exam (General Training or Academic) candidates are taking. English language teachers are often of the opinion that the General Training exam is easier than that Academic IELTS although the fact is that test content is slightly different rather than the constructs or areas of language proficiency that are being evaluated. This points to the need for a wide variety of stakeholders to gain greater understanding of what is being assessed by high-stakes language tests.

1.5.2 Claims made by the test developers

The apparent increase in the use of language tests for non-academic purposes in the New Zealand context may now necessitate an examination of the validity of using language tests scores for immigration, professional registration and migration. As Kane (2015) stated "To validate a proposed interpretation, it is necessary to be clear about what is being claimed" (p.4). For example, the IELTS website (ielts, 2019) markets the language test for education, migration or work and states that it is the English language test that opens doors.

The IELTS partners (the British Council, IDP: IELTS Australia and Cambridge Assessment English) claim that they administered 3.5 million tests in 2018 in 140 countries and that the test is now a household name in many countries (ielts, 2019). The test developers (ielts, 2019) promote the use of the test for the growing number of business and professional positions where English is needed. Further, IELTS claims to be fair to all so the consequences of test score interpretation and use should be beneficial for New Zealand employers, and for New Zealand society in general.

This research study will analyse the perceptions of test takers who live and work in New Zealand, together with the perceptions of New Zealand employers regarding the following claims:

- IELTS is a household name. It is the most widely used test for migration to Australia, Canada, New Zealand and the United Kingdom (ielts, 2019).

- IELTS test scores are a reliable gauge of test takers' ability to communicate in English (ielts, 2019).
- Organisations around the world rely on IELTS to help them select the right people. A wide range of other employers from sectors such as finance, government, construction, energy, aviation and tourism also require IELTS (ielts,2019).
- IELTS ensures quality and fairness. IELTS is recognised and trusted by individuals and organisations worldwide for its fairness, reliability, and high-quality standards (ielts, 2019).

Employers and decision makers (test users) who use the results of test to make decisions about test takers need to rely on language assessment systems, the claims they make and to feel the appropriate test results can reliably mean a person is competent in a language (Kane, 2013). Immigration authorities, professional workplace bodies and employers want to be able to apply language assessment wisely to make the best use of human resources in a global market. However, greater public debate between test developers and test users (Sen, 2009) may be needed to help these professionals with language assessment literacy (LAL) and thereby help them understand, use and interpret test scores more reliably.

1.6 Native speaker test takers

1.6.1 Definition of native speakers of English

Although the definition of what constitutes a native-speaker of English (or any other language) is contestable, my perception of this term is similar to the definition put forward by Ortega (2009) who explained that a native speaker has a language/languages as their mother tongue, first language or L1 when they have acquired the language/languages implicitly from birth up until the age of four or five years. On the other hand, non-native speakers of a language such as English are people with English as an additional language (EALs), second language (L2) learners who often learn the language explicitly while others may be multi-language learners. The acquisition of a language is complicated, and some young users of English may be simultaneous or sequential L2 bilinguals (Ortega, 2009). Nonetheless, there are many IELTS test takers who identify as having English as their first language, according to the IELTS statistics (ielts, 2019). Kunnan (2018) provides strong examples of how language tests have been used in unethical and unfair ways throughout history. It is, therefore, pertinent to question if the use of test score results has beneficial consequences for native speakers, as well as non-native speakers of English. Immigration changes have led to the increasing need for native speakers of English to take the IELTS test in New Zealand. The IELTS mobile website (ielts,

2019) states that the exam is fair to all, avoids cultural bias and accepts all varieties of native-speaker English so it is necessary to evaluate if this is, indeed, true.

One of the issues discussed by researchers in language testing is how native-speaker proficiency can be defined since native speakers of any language vary greatly in their language competence across the four skills (reading, writing, listening and speaking). For this reason, Hulstijn (2015) and Kunnan (2018) questioned whether it is necessary, or even desirable, for migrants to attain high levels of competence in English. Stricker (2004) pointed out, native speakers of English do not necessarily gain high band scores. In my position as an IELTS examiner, I am aware that native speakers of English do not automatically score 8.0, leave alone 9.0 in the Speaking and Writing modules of the exam. Indeed, one of the participants in Smith and Haslett's (2007) research claimed that "IELTS 7 could not be achieved by many current New Zealand students" (p.35). As there is not one standard version of English, it could be argued that IELTS needs to take into account world accents of English and modern writing conventions, which would be less discriminatory against 'Indian English' or 'Chinese English'. This is especially the case in today's world where global citizens compete for employment and migration for work purposes has become easier and more common (Ryan, 2016). Caines et al., (2014) pointed out that high-stakes exams are increasingly being used for immigration, personal registration and higher education but that global populations are diverse, so these types of exams should take into consideration their varying abilities to perform in the test.

Notwithstanding the abilities that are necessary to answer IELTS reading questions, the fact remains that even university lecturers can find IELTS reading tasks difficult to answer (Moore, Morton & Price, 2011). A surprising result was reported by Smith and Haslett (2007) when, in 2002, a Sunday newspaper in New Zealand engaged local residents (a New Zealand housewife, a New Zealand student, an Indian born receptionist and two multi-millionaire businessmen who had been resident in New Zealand for ten years) in taking a mock IELTS test. The first three participants completed the test but were physically drained, whereas the businessmen walked out (Smith & Haslett, 2007). Evidence of the likely band scores received by native speakers and non-native speakers of English, together with their perspectives on the relevance of the IELTS test for on the job competence may help elucidate if test scores are being used in valid and ethical ways. Further, the test score results from any language test for entry into a country cannot be deemed to be used ethically if non-native speakers of a language are required to gain higher scores than the average native speaker in a country.

1.7 Theoretical perspective

Issues related to migration, employment and social inclusion invariably encompass elements of social constructionism; migrants with English as a second language, their employers and colleagues are co-constructing an awareness of migrants' discourse competence and the language skills that are needed in the workplace. For this reason, I have used a qualitative approach since people taking IELTS, their employers and the institutional test users interpret the use of IELTS for the workplace in various ways. By taking this approach, the aim was to gain a rich narrative that reflects how each group constructs their awareness of high stakes language tests and how social interaction leads to shared perspectives which can change over time as described in Chapter 2. The myriad of perspectives which emerged were analysed using a hermeneutic approach to data analysis. The intention was to shed light on the symbolic meanings that groups of people (test takers and test users) attach to the use of language assessment for migration/work purposes and, hence, its validity for a variety of workers in the New Zealand context.

This study will examine the viewpoints of test takers (native and non-native speakers) as well as native speaker employers (test users) regarding the content and constructs of a high-stakes language test and how the test compares with the type of language used in the modern workplace. To assess the validity and ethical use of a high-stakes language test in New Zealand, an argument-based validation analysis was selected. This is because research into test validity has moved from the psychometric evaluation of individual traits and test construct reliability (Cronbach, 1988; Messick, 1989) to more encompassing studies into the evaluation of tests as social entities with intended and unintended consequences (Bachman & Palmer, 2010; Chapelle, Enright & Jamieson, 2008; Kane, 2013; Messick, 1989). Early argument-based validation research focused primarily on the development of tests at the initial stages. Yet, there has been an increased call for more in-depth studies into what test stakeholders infer about a test taker's competence in relation to their test score results and the wider social, economic and political implications of using high-stakes language testing (Chalhoub-Deville, 2016; Hall, 2010; Kunnan, 2018; Ricento, 2014; Shohamy, 2001; 2007).

1.8 Purpose and significance

The primary purpose of this study was to construct and investigate a validity argument approach for the ongoing debate about the validity and ethics of using high-stakes language tests as gatekeeping devices for immigration and professional registration purposes.

According to Kunnan (2018), language assessments should meet the needs of the various stakeholder groups as fairly and as beneficially as possible, while Knoch and Macqueen (2020) claim that language assessments which are used for professional purposes should be trustworthy because they inhabit policies with great societal power.

This research study is unique because it has included the perspectives of native speakers of English towards a test for non-native speakers of English and analysed which competencies are valued by the people who employ migrants to New Zealand. The perspectives of test-takers (native and non-native speakers of English) who are in employment and native speaker test users are crucial to the validity of tests. While earlier studies have concentrated on the validity of tests such as IELTS for academic success in tertiary studies, or for employment in the health and education sectors, there has not been much research into other employment areas apart from the investigation by Moore, Morton, Hall & Wallis (2015) into literacy practices in the professional workplace compared with IELTS reading and writing tasks. My research has gone one step further by analysing the viewpoints of native speakers test takers and native speakers who employ migrants.

Additionally, this study responds to the call for more research into the social dimension of language testing at the micro and macro levels by McNamara & Roever (2006). The micro level includes test takers' ability to use language in the domain for which the test is intended while the macro level relates to the use of language testing in policies which use language test results as gatekeeping devices. Other researchers have noted the need to investigate the social, economic and political implications of policies which use language testing results (Chalhoub-Deville, 2016; Hall, 2010; Spolsky, 2008).

From a theoretical perspective, this study contributes to research which uses an argument-based approach to test validation. Although this approach was utilised by Kane (2013) for educational testing in general and for language testing by Bachman and Palmer (2010), Chapelle (2011), and Chapelle et al. (2008), there are still not many studies or theses which utilise this approach.

From a practical perspective, while the focus was on the linguistic needs associated with workplace competence in New Zealand, aspects related to the changing use of language in the workplace (the importance of basic language cognition as opposed to higher language cognition) may be generalised to other contexts.

1.9 Research questions

The main question driving this study was 'To what extent are language test scores used in valid and ethical ways for the purposes of immigration, employment and professional registration in a New Zealand context?'

To assess test validity, it was necessary to consider if a test such as IELTS has a clear meaning for stakeholders (domain inference). This relates to the claim made by the test developers that IELTS is a household name and the most widely used test for migration to Australia, Canada, New Zealand and the United Kingdom. The research sub-question used to analyse the research participants' perceptions regarding the relevance of the test was:

> RQ1. How well do test takers and test users perceive the relevance of the assessment and the assessment tasks to the New Zealand domain?

To investigate if the test is being used ethically requires an investigation into whether employers can make appropriate decisions about test takers' linguistic competence (both native and non-native speakers) and their subsequent employability (decision inference) based on the test results. This relates to the test developers' claim that IELTS test scores are a reliable gauge of test takers' ability to communicate in English. The research sub-question used to analyse the research participants perceptions in this regard was:

> RQ2. How well do test users feel they can interpret and use test score results?

Since IELTS now claims organisations around the world rely on IELTS to help them select the right people in a wide range of sectors such as finance, government, construction, energy, aviation and tourism, the perceptions of employees and employers in similar employment sectors were used to answer the following research sub-question:

> RQ3. How well do test takers and test users feel that the IELTS tasks and the constructs being tested reflect communicative language used in the New Zealand workplace?

In addition, an examination of the participants' perceptions of the consequences for test takers, employers and New Zealand society (consequences inference) will help elucidate the test developers' claim that IELTS is fair to all and can be trusted by individuals and organisations worldwide. Kunnan (2018) stated that tests ought to be fair to test takers by being meaningful in what is meant to be assessed and having beneficial consequences for test takers. The following research sub-question was used to analyse the research data:

RQ4. What do test takers and test users perceive to be the consequences of test use for the New Zealand context?

1.10 Terminology

While I have attempted to keep acronyms and terminology to a minimum, a glossary of any terminology used in this thesis is provided on page 12.

1.11 Organisation of thesis

This chapter has given an overview of the research context, the catalysts for this study and introduced the research question as well as the four sub questions. In the next chapter (Chapter 2), previous research into high-stakes language testing for the workplace, issues for test taker and test users together with issues surrounding test score interpretation will be discussed. This will be complemented by a presentation of previous research which highlight issues surrounding the testing of English in a multi-cultural, global world: a description of world Englishes, the concept of higher language cognition versus basic language cognition (Cummins, 1984; Hulstijn, 2011, 2015), alternative ways of testing and the need for language assessment literacy. Finally, chapter 2 will focus on developments in validity and validation studies and will introduce the Inference Use Argument (IUA) approach to test validity, together with the call for research into test consequences. The research methodology and design will be outlined and justified in Chapter 3. Chapter 4 will present the perspectives of the research participants at an individual (micro) level in order to present their views on the relevance of IELTS assessment and assessment tasks for the NZ domain (RQ1) and how well they feel they can interpret and use test score results (RQ2). Chapter 5 will look at language used at the group or workplace level to explore how well test takers and test users feel that the IELTS tasks and the constructs being tested reflect communicative language used in the New Zealand workplace (RQ3). Chapter 6 addresses the research participants' perceptions of the social, economic, societal and political consequences (macro level) for New Zealand (RQ4). Chapter 7 discusses the implications of language assessment using domain and extrapolation, inferences to evaluate the validity of IELTS while the decision and consequence inferences will help illuminate if IELTS is being used in an ethical way. This chapter will conclude with a discussion of the implications for test takers, test users and policy makers when using test score results.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction to the literature review

This review of literature will describe high-stakes language assessments and how the reasons for taking them are changing. At the same time, the way language is being used in the contemporary workplace is changing, so one question that arises is: are the changes in language use in the workplace convergent or divergent with the way in which language is assessed by language testing systems?

This chapter will firstly review previous literature related to high-stakes language tests, the IELTS test in particular and previous investigations of the validity of IELTS for academic and non-academic purposes. Previous research has mostly analysed the skills language tests are assessing compared with the skills that are needed for a person to be viewed as competent in academia, although some researchers have investigated language skills assessed in language tests compared with those needed in the workplace (Moore et al., 2015). Secondly, there will be an analysis of research into the perspectives of language test stakeholders. To this end, research that analyses the validity of high stakes language tests will be described and how three different test stakeholders perceive test validity: namely, test developers, test takers and test users or decision makers (Caines et al., 2014; Kane, 2013). This will be complemented by a review of English language use and language test use in a multicultural, global environment. This is because language tests should be reflecting language use in society where English as a lingua franca, English as a world or international language is often used for communication in the workplace. For this reason, literature that looks at test validity in light of the debate about the varieties of English that are used to communicate in a global environment will be discussed (Brown 2014, Hulstijn 2011, 2015; Jenkins 2009). The question of whether basic language cognition or higher language cognition is practised in today's workplace by L1 and L2 speakers of English will be probed to ascertain if the language tested by language assessment is consistent with language use in society. This will lead to a section on possible changes to the current language assessment landscape that have been suggested in previous research literature. This includes using alternative forms of tests, sharing language assessment decision making between language testing experts and domain experts, together with more recent discourse that addresses the need to promote language assessment literacy education (Chalhoub-Deville, 2016; Lo Bianco, 2017) to a wider audience.

Finally, previous literature which explores new developments in assessment validation research will be introduced since recent validation arguments are not so concerned with the internal quality of the test but more with the use of test scores, the interpretation of test scores and the intended and unintended consequences of test use in society. For this reason, discourse that addresses the fairness and justice of educational assessment (Kunnan,2018), together with the need for Critical Language Testing (Shohamy, 2001) and an analysis of the social impact of assessments (Chalhoub-Deville, 2016) will be presented.

2.1 The use of high stakes language tests for immigration

As outlined in Chapter 1, high stakes language testing systems used globally include the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL, now commonly administered as an Internet-based test, the iBT TOEFL) although the Pearson Language Test (PET) is a newer computer-based test that is becoming popular. IELTS was taken by more than three million people world-wide in 2018. It is the English language assessment that is most commonly used in the New Zealand context as a prerequisite for academic study and is cited by Immigration New Zealand (INZ, 2018) as being one of the tests that are required for permanent residence in the skilled migrant category for principal applicants, their spouses and children.

2.1.1 Changes in the use of high stakes language tests – from academic to professional registration and business purposes

IELTS was initially developed in 1989 to test international students from non-English speaking backgrounds who wished to undertake academic courses in English at tertiary institutes (Merrifield, 2008; 2011; Moore et al., 2011; Rumsey et al, 2016). Eventually, immigration authorities began to use this testing system as a means to assess migrants' suitability to successfully integrate into English speaking countries (Merrifield, 2008; 2011). A more recent development has been the use of this high-stakes language test for professional registration and business purposes. Merrifield (2008) states that between 2004 and 2006, the number of candidates taking IELTS for professional purposes rose by 100% in Australia and by "somewhere in the vicinity of 500% in the USA" (p.13). In New Zealand, people seeking to work in a wide variety of jobs, ranging from the builders, chefs and sheep shearers to lecturers and lawyers, are currently required to take this exam. Yet, as Caines et al. (2014) point out, high-stakes exams can "have lasting and sometimes adverse consequences for test takers" (p. 6). This implies that it is necessary to research the positive and negative consequences of high-stakes language tests used for business, professional and migration purposes in New Zealand.

2.1.2 Previous research into the IELTS test for professional registration and business employment purposes

Most research into high-stakes language testing has investigated test impact on academic progress. However, recent articles commissioned and published by IELTS have investigated the relevance of IELTS test materials in the four macroskills (reading, writing, speaking and listening) in the workplace, but predominantly for the healthcare sectors: for overseas-trained nurses in the U.K. (Sedgewick, Garner & Vicente-Macia, 2016) and for healthcare workers and early childhood teachers in Australia (Gribble et al., 2016). In New Zealand, overseas-trained nurses are frequently employed in the burgeoning retirement home industry, but they may often struggle to gain the required band score of 7.0 across the four skills to register as nurses (Gribble et al., 2016). While Read and Wette (2009) undertook valuable research into the attitudes of overseas-trained health care professionals towards IELTS and the Occupational English Test (OET) in a New Zealand context, there has been little research into the relevance of IELTS as a testing mechanism for other spheres of employment such as the trades, engineering and IT staff, included in skilled migration shortage lists.

Research into the validity of IELTS in business and professional domains was conducted by Moore et al. (2015), but this was in an Australian context. Moore et al. (2015) found that the type of language needed in the workplace is simple, clear and easily understood by clients. Similarly, Murray et al. (2014), analysed the situation for teachers in Australia and New Zealand but with the current teacher shortage in New Zealand, it would seem apposite to focus on the perspectives of New Zealand teachers and the people who employ them anew.

2.2 Issues related to the validity of high-stakes language tests

2.2.1 The validity of language tests for academic study – comparison of constructs

Although it may be assumed that higher language proficiency would be needed to pursue doctoral studies compared with general migration, the required band score to study a PhD in New Zealand is 6.5 while principal migrants for immigration also need a 6.5 band score. Moore et al. (2011) researched the constructs being tested by IELTS in the reading module and compared them with the skills needed for academic study. This research analysed the perspectives of university lecturers and found that IELTS tasks did not reflect the types of reading skills needed at university. Lecturers at two tertiary institutions compared their course content with typical IELTS reading tasks. This study was limited because the focus was only on the reading section of IELTS. However, the concept of construct validity was succinctly

outlined and described as being “related to those abilities it is thought readers need to possess in order to handle the demands of the target language domain” (Moore et al., 2011, p.6). It was found that the IELTS reading tasks required a more literal and local reading process as opposed to an interpretive and global level of engagement, which is expected at university. Although some IELTS questions entailed an awareness of textual meaning at a paragraph level, most needed an interpretation of language at the sentence level. When the lecturers in the study by Moore et al. (2011) were asked to assess the IELTS reading tasks and reading processes, it is interesting to note that the communications lecturer stated “I found the science-based articles and items quite complicated actually. If I had to answer questions about the science, I'd have to go back and read it twice” (Moore et al., 2011, p. 60). If a native speaker lecturer found an IELTS reading task difficult, this points to task complexity in the IELTS exam. However, as the authors (Moore et al., 2011) state, IELTS tasks are more aligned with a positivist epistemological stance as opposed to a constructivist one, meaning that scientifically oriented test takers are likely to fare better in the reading section. It was apparent that the university lecturers in Moore et al.’s (2011) study had not been aware of the type of skills tested by IELTS, which suggests that the IELTS partner institutions (British Council, IDP Australia and the University of Cambridge) need to put more effort into informing people who rely on the validity of IELTS as regards to what is actually being tested.

Similarly, Pilcher and Richards (2017) interviewed lecturers in six academic fields to discover if IELTS tasks reflected the constructs needed to succeed at university. Although Pilcher and Richards (2017) discuss the four skills tested by IELTS, this is merely a brief overview to point out that the key constructs assessed in an IELTS test are grammatical accuracy, accurate spelling, spontaneity and flexibility in all four skills. This study (Pilcher & Richards, 2017) ultimately concentrated on lexis and specifically how lexis is appropriated and used differently in six academic fields. Interestingly, none of the lecturers expected ‘perfect’ English from students whose first language was not English, and the language valued in each subject area was different; some lecturers valued numbers or technical jargon while others noted that psychological and emotive language were essential in their academic fields. Crucially, Pilcher and Richards (2017) highlighted that the context in which language is used plays the most important role in most subject areas whereas IELTS tests de-contextualised, neutralised language and can therefore not necessarily be relied upon to determine the language skills needed for academic success at university.

Neither Moore et al. (2011) nor Pilcher and Richards (2017) mention the levels of English proficiency expected of native English-speaking students, and they do not compare the four skills in the IELTS test with the competence needed in these four skills across different academic fields of study (the target language use domains).

2.2.2 The validity of language tests for the professional workplace – comparison of constructs

As regards workplace competence, Moore et al. (2015) looked at the constructs being assessed in the IELTS reading and writing modules (both academic and general training), comparing them with the skills commonly required in nine professional domains (accounting, management, law, engineering, science, IT, health, education and media). In Moore et al.'s (2015) study, some of the skills necessary to succeed at IELTS were found to be useful but both Moore et al.'s (2011 and 2015) pieces of research found that IELTS tasks do not adequately cover the skills utilized in either academic study or the professional workplace.

Looking at workplace communication, this is characterized by the need to read and produce emails in a brief concise fashion using plain English (Moore et al., 2015) whereas a high level IELTS candidate would be expected to produce oral and written output using a wide range of syntactical and lexical structures accurately and flexibly. In the later study by Moore et al. (2015), the authors highlighted that, as far as writing is concerned, the Task One (Academic) and Task Two questions (Academic and General Training) do not have a specified audience, which is in direct contrast to workplace correspondence. While the Task One (General Training) is a letter, which is similar to email, and the candidates are expected to change the tone of their writing in relation to a specified audience, the task is more of a consumer type transaction (complaining to a hotel manager, explaining why the writer cannot attend an exam) rather than service to clients. As regards reading, according to Moore et al. (2015), graduates in the workplace need to be able to extract the key points of an email or manual and then determine how to progress to the next stage. In this respect, the General Training writing module is more appropriate because it is more expository and deontic in nature (identifying whether something is going to be done) as opposed to epistemic (identifying whether something is true), the latter being more useful for academic purposes (Moore et al., 2015).

While it may be the case that an exam such as IELTS cannot cover lexical and syntactical conventions across a range of professional domains, Moore at al., (2015) claim that generic

formal or semi-formal skills useful for workplace interaction include making requests to colleagues, reporting on activities completed and getting clarification from superiors. The authors recommend that Task One (GT) be more oriented towards client issues and possible actions and that Task Two questions be adapted to a problem-solution justification framework, which is more common in the workplace.

2.2.3 Academic versus General Training modules of IELTS for immigration, employment and professional registration

Language tests can be designed for specific purposes, and generally these tests can be divided into those that test language for academic purposes or language for vocational or professional domains (Basturkmen & Elder, 2004; Douglas, 2000; Knoch & Macqueen, 2016). IELTS is not a specific test; rather both the Academic and General modules can be placed at the general end of the general – specific assessment specificity continuum (Knoch & Macqueen, 2016). Test users, however, are not usually aware that the same speaking and listening modules of IELTS are used for both the Academic and General Training versions of the test. For this reason, Aryadoust (2013) questioned the claims made by IELTS that the listening module can effectively test constructs used in both academic and non-academic spheres.

According to Merrifield (2008), test users in her study into the use of IELTS as an entry criterion for professional associations in Australia, New Zealand and the USA, had little knowledge of the difference between the General and Academic modules. Although the IELTS partners have changed the content of the reading and writing tasks (GT) to better reflect the fact that test takers are increasingly taking the exam for professional reasons, they have not altered the test constructs, or the rhetorical styles needed to be used by candidates. Ultimately, Moore et al., (2015) call for further research into test users' perceptions of the General versus Academic Training modules and the skills they identify as being useful for workplace competence so that the training undertaken by students who have English as an additional language can be beneficial for the students and society in general. The skills that were not researched in these studies (Moore et al., 2011; Moore et al., 2015) were those of listening and speaking, which could be arguably considered as being more relevant for modern day work-based competence. In my research, some participants reflected on the differences between the Academic and General Training modules of IELTS, but participants predominantly commented on the skills and language abilities which they perceived to be relevant in the New Zealand workplace.

2.2.4 Test consequences for migration, employment and professional registration

The issue of predictive validity was mentioned by Elder and O'Loughlin (2003) when researching typical band score gains of international students taking IELTS courses in Australia and New Zealand. They (Elder & O'Loughlin, 2003) noted that the role of IELTS is influential in determining test takers' future opportunities in life; therefore, it is "of vital importance that research is conducted into the uses which are made of IELTS test scores and the meanings which are ascribed to them" (Elder & O'Loughlin, 2003, p. 208). This may be more imperative in today's global environment where IELTS is used, not only to enter academic institutions but increasingly for migration and professional registration. There are numerous ways to address the validity of high-stakes exams but as Kane (2013) claims, for a test to be seen as being valid, there needs to be clear information about how the test scores are interpreted and used in society. Particularly when tests are used "to predict future outcomes (in an employment- or placement-testing context), evidence indicating that the predictions are accurate is called for" (Kane, 2013, p. 3). Similarly, Caines et al. (2014) consider the viewpoints of test takers and test users as being essential to validity arguments since test scores can be misinterpreted by them. It is important, therefore, that both these groups understand exactly what is being tested and how that interfaces with real life and the future use of language in context. Employers' perspectives in different professions regarding language tests versus workplace competence have not been covered in depth in previous research.

2.3 Test takers' and test users' perspectives regarding high stakes language tests

The users of high-stakes language tests can be divided into three groups: test developers, test takers and test users (Caines et al., 2014). The validity of the exam may be perceived differently by each group. Test developers (IELTS, in this instance) see validity as being linked to how accurately the test measures aspects of language across a wide range of candidates world-wide while test takers have a 'contest perspective' and naturally want the highest score possible (Caines et al., 2014) even though the majority of native speakers of English cannot gain a band score of 9.0 across all modules (ielts, 2019). To a certain extent, test takers also want the scores to predict their ability to succeed socially and economically. Likewise, test users (the immigration authorities or professional bodies), are seeking predictive validity in so far as they want reassurance that they can predict that a person with an IELTS score of 6.5, for instance, has the requisite skills to be minimally competent 'on the job' or in society. The issue of integration was also raised by Gibb (2015), who claimed that standardized language tests exacerbate the power imbalance felt by newcomers to a country. It is often not acknowledged

that migrants with an additional language bring with them existing knowledge and skills, as well as cross-cultural competence (Sawyer & Singh, 2012). The language that they use may not be 6.5 IELTS or above, but it may be perfectly adequate to be seen as comprehensible for the workplace. For this reason, Gibb (2015) advocates exams which assess how migrants collaborate and communicate at work.

The consequence of high stakes tests such as IELTS may be categorised as being negative if test results lack predictive validity as regards the ability to communicate effectively in a workplace context. Gribble et al. (2016), discuss instances when nurses were already seen as competent on the job, were well-liked by their colleagues and appreciated in the workplace; yet, they found it difficult to obtain a 7.0 band score in IELTS. This is illustrated in a quote from their research from a manager working at an aged-care facility:

> "We have a number of people, especially at the diploma level, who got through their diploma or degree, but when they sat the IELTS, they couldn't get through. Now I'll cite an example of a woman that's been with us for a number of years, she works on night duty, her manager heaps praise on her, she does a good job. They have no difficulty with her communication skills. She sat the IELTS three times and she's failed each time – she just can't get through it. And that's not a one-off either, I've had a number of managers come to me and say is there anything else we can do to support this person, because we want to keep them on" (p.39).

Caines et al., (2014) and Kane (2013) allege the viewpoints of test users and test takers are essential to assess test validity and this formed the basis of my research.

2.3.1 Issues for test takers

As they sit the test, it would seem reasonable to assume that test takers fully comprehend what IELTS measures and how this facilitates or impedes their trajectories towards work or social inclusion. Yet, this may not always be the case because test takers do not necessarily question the testing process, believing tests are credible and infallible (Kunnan, 2018). The qualitative approach taken by Roshid and Chowdhury (2013) in Australia investigated Bangladeshi migrants with Master's level degrees who had obtained a wide range of jobs, from cleaners to university lecturers, noting their IELTS band scores, together with their perspectives on the importance of English for employment. The findings showed a link between English proficiency and the ability to obtain work. However, what was not noted by Roshid and Chowdhury (2013) was that the two participants with employment which was

commensurate with their prior experience and qualifications, had actually gained Australian qualifications and, in all likelihood, socio-pragmatic competence while studying in the country. In addition, participants pointed to the importance of 'knowing people' and 'proving yourself capable' on the job as being equally, if not more, beneficial than language skills.

Similarly, Rumsey et al., (2016) researched the mobility of overseas-trained healthcare workers but as their research progressed, they discovered that the need to take IELTS was more of an issue for their participants than moving to Australia. Indeed, one participant had worked in Australia for twenty years but still needed to take IELTS six times to gain the requisite band score. One of the most pertinent quotes came from a specialist who stated:

> "I've been working in very senior positions. I've been in emergency departments for 4 years. I have no trouble with English but the IELTS is very complex and doesn't test your skills in English it tests your secretarial skills" (Rumsey et al., 2016, p. 101).

This is a narrative that is re-iterated by many researchers (Gribble et al., 2016; Read & Wette, 2009; Rumsey et al., 2016; Sawyer & Singh, 2012; Smith & Haslett, 2007). One of the key findings of Rumsey et al., (2016) was that the healthcare professionals in Australia had a negative perception of IELTS to the extent that they believed it could be a money-making exercise. The Bangladeshi migrants interviewed by Roshid and Chowdhury (2013) felt that becoming competent in listening and speaking was crucial for migration and employment in Australia. Yet, becoming 'native speaker-like' was not deemed necessary (Roshid & Chowdhury, 2013). In general, in finance, healthcare and teaching circles, clear communication was deemed necessary for health and safety reasons (Rumsey et al., 2016). More importantly, it was reported that non-native employees need discourse and pragmatic skills in order to interact with the general public who represent a variety of ages, cultural and educational backgrounds rather than academic error-free English (O'Loughlin, 2006; Rumsey et al., 2016).

Teachers are expected to have excellent language skills (Murray et al., 2014). However, Al-Issa, Al-Bulushi and Al-Zadjali (2016) discovered that it was difficult for non-native student teachers of English to gain a band score of 6.0 IELTS. Al-Issa et al. (2016) researched the relationship between IELTS testing and the educational practices in Oman with the result being that some students felt they were underprivileged compared with IELTS candidates from other countries. Elder and Elder and Brown (Elder, 1993; Elder & Brown, 1997) commented on the fact that subject specialist teachers assess competence on the job as being governed by "sensitivity to

audience, interactive skill and personal style" (p.88) whereas language experts focus primarily on the quality of discrete aspects of language. This highlights the crucial distinction between language competence for the workplace and traditional linguistic proficiency as measured by IELTS. Furthermore, Murray et al. (2014) called for research in the teaching profession which analysed the connections between language competence and how effective teachers perceive themselves to be in the workplace.

2.3.2 Native speaker test takers

The construct of the native speaker is hard to define (Hulstijn, 2015). Nonetheless, native speakers generally learn their mother tongue in their formative years and all (if not cognitively impaired) implicitly gain high levels of basic language competence (Loewen, 2015; Ortega, 2013).

Native-speaker candidates may be expected by decision makers to gain higher scores than non-native speaker candidates in a language test. Yet, even native-speaker university lecturers reported that they found IELTS reading tasks difficult to answer (Moore et al., 2011). A surprising result was reported by Smith and Haslett (2007) when, in 2002, a Sunday newspaper in New Zealand engaged local residents in taking a mock IELTS test. They all found the test too onerous. In the USA, both native and non-native speakers of English who want to have their veterinary qualifications validated, are required to take an English assessment for PAVE (Program for the Assessment of Veterinary Education Equivalence) with candidates raised and educated in the U.S. and Canada not necessarily out-performing NESB (non-English speaking background) students (Merrifield, 2008).

An analysis of the 2017 IELTS statistical data (ielts, 2019) reveals that speakers of English as a first language who took the Academic test were outperformed by candidates who have German as their first language. Candidates who identified themselves as having English as their first language attained a mean overall score of 6.92, compared with 7.41 for German speaking candidates (the mean band scores for English speakers were Listening 7.21, Reading 6.71, Writing 6.35, Speaking 7.14 compared with German speakers who scored Listening 7.8, Reading 7.55, Writing 6.62, Speaking 7.39). Candidates with Filipino and Polish as their first language also attained very similar scores to the English speakers with a mean overall score of 6.88 (Filipino) and 7.03 (Polish). It must be noted that the candidates themselves choose their first language from a wide variety of language options, so it is impossible to verify if English is actually their first language. Nonetheless, professional associations and immigration

authorities may be surprised to discover that a mere two percent of English speakers attained a band 9.0 in the General Training test in 2017 while for the Academic test the percentage of English speakers who received a band 9.0 was zero.

McNamara and Roever (2006) claim that it is unethical to measure language proficiency against idealised monolingual speakers of a language. It may be deemed unethical to apply standards to non-native speakers of a language that are unattainable, even by those presumed to be competent in a language. Indeed, Stricker (2004) compared the language output of ESL students and native speakers of English who took the TOEFL iBT and GRE (Graduate Record Examination) exams in five states in the US between 1999 and 2000. He alleged that the native speakers did not "perform perfectly" (p. 160).

There has been little or no empirical research that I have found in the Language Testing Journal or in the IELTS research reports that looks into the perspectives of native speakers of English, although as an examiner I have tested candidates from the Canada, New Zealand and the United Kingdom, that is from the inner circle of countries (Brown, 2014) where English is considered a first language. Stricker (2004) concluded that native speakers vary in linguistic, educational and cognitive skills and are, therefore, not necessarily the "ultimate criterion group" (Stricker, 2004, p. 137). There is also scant research that has analysed the perspectives of test users towards language tests and language proficiency.

2.3.3 Issues related to test users' understanding and acceptance of different band scores for immigration, professional registration and employment

Although many professional associations require an overall IELTS band score of 6.5 or 7.0, there seems to be little consistency across professions. The current IELTS website (ielts, 2019) indicates that band scores of between 3.0 and 8.0 are deemed appropriate by a variety of employers and professional institutions worldwide. The differences between these scores is very large, ranging from beginner level users of English with low communicative competence to advanced communicators, who are able to relay information accurately, professionally and in-depth in a similar fashion to an 'educated' native speaker of English.

Likewise, there is confusion among associations who accept IELTS. Registering bodies often appear to be puzzled about minimum standards. They are also not aware that empirical research exists which identifies minimum standards. For example, when Elder (1993) was asked about the optimum IELTS threshold for entry to teacher education, she replied that there were many factors which affected the success of learners who wanted to be teachers.

These include knowledge of the subject, the classroom culture and pedagogical awareness. As a result, any score above global bands 4.5 and listening band 5.5 may be appropriate in making initial student selection.

The panellists in the study by O'Neill, Buckendahl, Plake and Taylor (2007) which looked at acceptable scores for the nursing profession, recommended IELTS scores of between 5.5 and 5.6 for Speaking, between 5.2 and 5.3 for Task One Writing and 5.4 for Task Two Writing whereas a minimum of 6.5 was deemed appropriate for Reading and Listening. These recommendations are relevant because the panellists included not only nurse educators and regulators but also NESB (non-English speaking background) nurses and nurses who regularly worked with NESB patients. This panel of experts, therefore, included people with a knowledge of language acquisition and communicative competence as well as professional content knowledge. Yet, when O'Neill et al.'s (2007) research was published the prerequisites for nurses were 5.5 in the UK and 7.5 in New Zealand, pointing to a significant difference in the required minimum thresholds. Furthermore, Merrifield (2008; 2011) presented an overview of required band scores for immigration and as an entry criterion for professional associations in Australia, New Zealand, Canada, the UK and USA. She discovered that the majority of test takers and test users deemed it necessary to have the highest IELTS score (7.0 or 7.5) in the Speaking module, which is quite different to the recommendations in the study by O'Neill et al. (2007).

These discrepancies point to the fact that although test users want the exam to be reliable in order for them to be able to choose a person with the abilities to perform successfully in the workplace and to integrate in the wider community (Caines et al., 2014), there is insufficient awareness among them of what the band scores mean or even the difference between the IELTS Academic and General modules. Even associations who used IELTS as a benchmarking tool in Merrifield's (2008) study (the IMF, USA and Institute of International Education, USA), together with the Association of Chartered Certified Accountants, Australia and the Accounting and Finance Association of Australia and New Zealand were reluctant to participate in Merrifield's (2008) research because they were "not familiar enough" (p.7) with IELTS as a testing system. Likewise, university entrance administrators (O'Loughlin, 2013) and government policy makers (Pill & Harding, 2013) had little knowledge of what is involved in language testing. This impacts on the validity of IELTS for migration and employment purposes.

2.4 Test score interpretation by employers, immigration and professional bodies

Test users may make judgements about people who have taken high-stakes language exams but Caines et al. (2014), Kane (2013) and Gibb (2015) question how they interpret the test scores and wonder about the connection between test scores and the future use of language. For example, IELTS recommends that test users invest time and energy into researching the appropriateness of band scores for their workplace context (ielts, 2019). However, the web page, which is titled IELTS for work and is aimed at professional organisations, actually gives guidance for educational institutions. As for 'on the job' competence, Merrifield (2011) alleges that "none of the respondents were able to reference any in-depth research into the language skills required to operate effectively in the profession which they represented" (p. 43). This corroborates the findings from Merrifield's previous research (2008) which found little language assessment literacy amongst test users. IELTS does organise information sessions for test users to help them interpret the band scores but it appears that many would like more contact with IELTS; an information pack alone does not suffice. Some of the people interviewed in Merrifield's research (2008; 2011) stated that they would like to try the test themselves in order to understand the test and its purpose. Even if employers and professional organisations had sufficient time to dedicate to understanding what exactly is tested by IELTS, the people who work in immigration or for professional bodies are not linguists and find the differences between the scores difficult to understand (Merrifield, 2008). In fact, in the same article Merrifield (2008) went further, alleging that "the rating band descriptors for Speaking and Writing have only recently been made public, and stakeholders who are experts in the ELT field have had difficulty interpreting the scores" (p.28).

The healthcare sector has used IELTS for a long period of time (Merrifield, 2008; Read & Wette, 2009) and, for this reason, decision makers in this area are more aware of the language proficiency abilities that prospective employees are likely to have if they have attained a particular band score. However, Merrifield (2011) still reported that this awareness was often superficial; decision-makers did not know the difference between the General and Academic modules and could not say why IELTS had been chosen as a language assessment tool. According to Merrifield (2008; 2011) people who make decisions about whether to use IELTS as a testing system and understand the test results do not often remain in their roles for a long period. As a result, organisations frequently adopt the same band scores as other bodies (Merrifield, 2008; 2011), for example the New Zealand Nursing Council followed their counterparts in Australia when raising the IELTS requirements.

A similar situation can be seen in academic circles; according to Moore et al. (2011), university lecturers were unaware of what IELTS tests and assumed that a score predicts academic prowess rather than accuracy of language. Likewise, the school principals interviewed in the research by Murray et al. (2014) had very little knowledge of issues related to language proficiency and what exactly IELTS assesses. Some believed that it measured communicative competence in a professional context. These principals expected the non-native teachers they employed to obtain a high band 8.0 and to produce error-free written language. Yet, such a high band score may not result in teachers who are competent in the workplace (Elder 1993; Gribble et al., 2016; Murray et al., 2014). Elder, as far back as 1993, found that IELTS may be relied on to indicate performance in teacher education courses in the short term. In the long term, however, "the predictive power of IELTS diminishes because of improvements in English language ability and the nature of the language variables likely to affect teaching performance" (Elder, 1993, pp.78-80). As Murray et al. (2014) argue, English language competence does not lead to effective teaching practice.

The aspect of language testing that is often not sufficiently highlighted, and an aspect that is not understood by gatekeepers in general, is that IELTS is a snapshot of a test taker's English language proficiency on a particular day (Spolsky, 2008). According to Pilcher and Richards (2017), objectivists view language as being stable and immutable, but in reality language is social and context dependent. As Van Moere and Hanlon (2018) point out, there is a tendency to interpret scores from single administration tests as accurate indicators of a latent trait, but that this is not a reliable measure and should, therefore, be treated with caution. Cook, Brydges, Ginsburg and Hatala (2015) stated, in their research into validity arguments for assessment of doctors and surgeons, that in assessments it is common for a number to be generated, but that this number is of little value to test users since they do not fully comprehend what these numbers represent. It is therefore debatable if it is ethical to use test score results to validate a test taker's academic prowess or their ability to be competent communicators in English for either work purposes or social integration. Educational test results may vary each time a candidate takes a test (Bonk, 2018) and language competence cannot be described as a latent unidimensional trait (Pilcher & Richards, 2017; Spolsky, 2008; Van Moere & Hanlon, 2018) since language acquisition (for native and non-native speakers of a language) is dynamic.

2.5 The nature of English language test use in relation to world Englishes and the type of language cognition used in a multicultural global environment

2.5.1 World Englishes

In today's globalised environment, the varieties of English spoken have increased. As Brown (2014) claims, "the English native speaker norm is no longer sacred" (p.5). The term world Englishes (WE) encompasses *inner-circle* (from countries like the UK, USA, Australia)*, outer-circle* (India, Nigeria, Singapore) *and expanding-circle* (China, Japan, Russia) *Englishes.* One of the underlying beliefs of proponents of World Englishes is that the English language now belongs to everyone who uses it (Kachru, 1986). This begs the question: to what extent can, or should, language testing include a variety of Englishes? Native speaker English is not easy to define as there are different accents, morphosyntactic and lexical differences in the *inner-circle* countries (Brown, 2014; Rose & Syrbe, 2018). Furthermore, native speakers of a language have differing educational and professional backgrounds which result in exposure to varying amounts of their first language (Hulstijn, 2011). Therefore, questions that have arisen include: What should be the standard, who should judge it and does the standard meet the needs of test takers who want to use the language for intercultural communication? These issues are particularly relevant since language testing systems such as IELTS are marketing their tests "in countries or with populations for which they were originally not intended" (Chalhoub-Deville & Wigglesworth, 2005, p. 383).

When compared with a native-speaker 'norm', the speaker of English from India may unfairly be judged as being deficient because he/she is being compared not with a native-speaker 'norm' but an 'idealised native-speaker norm' (Davies, 2011). Yet, there are more non-native speakers of English than native speakers in the world. The speaking module of IELTS may, in fact, be measuring English proficiency against "unrealistic and irrelevant standards" (Kim, 2006, p. 32) as what matters most is intelligibility in a post-modern globalised environment where speakers of English need to shuttle between multilingual communities (Canagarajah, 2006). Today, speakers of English have the difficult task of needing to understand and acquire the local linguacultural English norms (in India, the UK or Australasia) as well as competence in a more global 'standard' variety. On top of this, the English language is not static; it changes over time and some native speakers have competence in one domain of target language use whereas others have competence in multiple domains.

This suggests that language testing systems should ideally be testing strategic competence in English as a lingua franca (Elder & Davies, 2006), world or international English. Also, they should assess performance as being acceptable or unacceptable in the future context of use and that the ideal should be the 'good' local rather than the expert native speaker. However, it is important to note that English as a lingua franca (ELF) or as an international language (Jenkins, 2009) has multiple varieties, which are ever changing, and not having a particular standard can do a disservice to test takers. Ironically, if English as a lingua franca was well documented, it would itself become a 'standard'. Additionally, test takers often prefer to learn and be tested in 'prestige' varieties of English from the *inner-circle* (Elder & Davies, 2006). High stakes international language testing systems have to take many variables into account, including construct validity, fairness, accountability and predictive validity. However, Brown (2014) alleges that the current focus on discrete item accuracy is useful for test developers because such tests are easier to develop, have stronger reliability and lend themselves to statistical reporting. This means that language testing systems test what they can measure but this may not be a useful measure of what test users want or need to understand.

2.5.2 Basic language cognition versus higher language cognition

Looking at varieties of English from another viewpoint, Cummins (1984) and Hulstijn (2011; 2015) posit that communicative competence and academic proficiency in a language are not identical and that academic language is not necessarily acquired by native speakers. A factor identified by Cummins (1984) as being important for teachers of bilingual or multilingual English Language Learners (ELLs) is the difference between basic interpersonal communication skills (BICS) and cognitive academic language proficiency (CALP). BICS is context-embedded and develops much earlier than CALP, partly because academic language proficiency is context reduced. Cummins wanted to highlight the fact that English language learners at secondary school may seem to have good speaking and listening communicative ability in English but that they need to receive support in academic language proficiency. What is not stressed enough by Cummins (1984) is that CALP is often not developed in native speakers of English and that academic language proficiency is most evident in the reading and writing sections of language assessments.

In a similar vein, Carroll and Bailey (2016) researched English language proficiency assessments in the USA for native and non-native students in the fifth grade. Carroll and Bailey (2016) used native speaker students as a comparison group, hypothesising that if high

performing native speaker students were rated as being non-proficient when they took the test, the test could not be considered valid. Carroll and Bailey (2016) used Kane's (2013) interpretation/use (IUA) argument and discovered that "The decision rule based on non-existent monolingual speakers proved to be too difficult for the real monolinguals in our study, even the highest achievers" (p. 46).

One of the few researchers to address the issue of native speakers' differing levels of language proficiency is Hulstijn (2011; 2015). In his article on language proficiency in native and non-native speakers of languages, Hulstijn (2011) distinguishes between basic and higher language cognition, as well as between core and peripheral elements of language proficiency. Essentially, Hulstijn (2011) argues that all native speakers of a language attain basic competence in their first language in the sense that they can implicitly and automatically use the most frequent lexical and morphosyntactic structures of the language in any communicative situation. Hulstijn (2011) defines basic language cognition thus:

> BLC pertains to (a) the largely implicit, unconscious knowledge in the domains of phonetics, prosody, phonology, morphology and syntax; (b) the largely explicit, conscious knowledge in the lexical domain (form-meaning mappings), in combination with (c) the automaticity with which these types of knowledge can be processed" (p. 230).

As Hulstijn (2011) emphasises, higher language cognition may be dependent on an adult native speakers' intellectual, educational, occupational, or leisure time backgrounds. Higher language cognition includes the ability to use low frequency lexical items and more complex morphosyntactic structures, and may, for this reason, be dependent on educational level. Further, native speakers are very competent in BLC in terms of listening and speaking skills (especially with prosody and phonology) but they vary greatly in HLC, particularly with reading and writing. According to Hulstijn (2011), such variations in language proficiency should be regarded as being on a continuum (rather than a dichotomy) with most native speakers differing in all four language skills. Interestingly, non-native speakers of English may attain high levels of HLC in their L2 but elements of BLC may still evade them. This is especially dependent on their L1 in the case of pronunciation and certain elements of morphosyntax (in the case of English for example, articles, personal pronouns, third person 's', and the present perfect aspect). Unfortunately, the ability for non-native speakers to gain high levels of

proficiency (lexically and morphosyntactically) in professional domains but not in basic home related domains is mentioned by Hulstijn (2011), but not covered adequately in his study.

The ultimate consequence of the differences between BLC and HLC is that native speakers of a language may find the reading and writing modules of an exam such as IELTS challenging. As Hulstijn (2011) contends "the higher levels of the competence scales will also be normally attainable by people with higher levels of education, or functioning in higher professions" (p. 240). Hulstijn compares BLC and HLC with second language assessment conducted on the CEFR (Common European Framework of Reference for Languages), noting the inequity of CEFR assessment because "In fact, many adult native speakers will never attain the highest CEFR levels [C1 and C2]. However, the authors [the CEFR authors] do not explain why this is so, not do they explicitly acknowledge that the C1 and C2 levels will generally not be attainable by L2 users with educational backgrounds other than higher education" (p. 241). For a comparison between CEFR and IELTS scores, see appendix R.

While Hulstijn (2011) argues that his definition of basic language proficiency is more encompassing than Cummins's BICS, he still acknowledges that more research needs to be undertaken into how individual differences, especially in metacognitive strategies (which Hulstijn describes as periphery skills), affect native-speakers' language proficiency. According to Hulstijn (2011) the main difference between his description of BLC and Cummins's BICS is that BLC "explicitly refers to the distinction between language reception and production" (p.233). However, this is confusing because he refers mostly to oral language production when using examples of the CEFR rating descriptors to highlight that attaining the highest score results (C2, C1 or B2) is a reflection of a candidate's educational or professional status. Although Hulstijn stated that reading and writing typically belong to HLC, there is not sufficient evidence in his article (2011) or book (2015) to warrant this claim. I contend that HLC is discernible in native speakers' use of English, not in language reception versus production, but rather in reading (receptive) and writing (productive) skills as opposed to listening and speaking skills.

Another area not covered adequately by Hulstijn (2011) is that of uneven profiles. He reports that one of the CEFR authors has stated that uneven profiles are common, but he is referring to differences in linguistic performance within one skill as opposed to across skills. In the IELTS testing centre that I am aligned with, it is indeed normal to have many uneven profiles that need to be rated a second, and sometimes third, time. IELTS refers to these as jagged profiles.

The prevalence of jagged profiles received in tests such as IELTS by native and non-native speakers of English is an area where more research is needed.

2.5.3 Possible changes: Alternative ways of testing

One of the key recommendations made by Brown (2014) is that language tests need to recognise the "multidimensional nature of language" (p.18) by using task-based items, portfolios, conferences and self-assessments, amongst others and employing analytic rubrics to score them. Although this may be difficult to implement, if testing systems move from assessing formal grammatical competence of the 'expert' user and "develop instruments that are sensitive to performance and pragmatics" (Canagarajah, 2006, p.229), they may better reflect multicultural workplaces. Task-based language testing could, in fact, assess language as it is used in daily life. The Gordon Commission on the Future of Testing (2013) states that tests may not be so universally applicable in a globalised society and that the emphasis should move to "learning in an increasingly complex world" (Caines et al., 2014, p. 10). Another factor for consideration noted in this article is that global populations are diverse, and any testing mechanism should take account of their varying abilities to perform in the test.

Brown (2014) believes that proponents of World Englishes and language testers are more at cross purposes rather than in disagreement. One main challenge is to define exactly what *comprehensibility* and *intelligibility* mean. In addition, it may be ideal if this were not a one-way exercise; in a modern multilingual/multicultural world, it is interesting to note that not only speakers of English as a second language, but also native speakers of English may need the ability to understand and respond to the varieties of English spoken in the *outer* and *expanding* circles.

The notion of equitability in testing both native speakers and non-native speakers has been addressed in the Netherlands where it has been proposed that the national language, Dutch, be assessed at secondary school in a framework similar to the CEFR. However, as Hulstijn (2011) points out, it has not necessarily been made clear that this new L1 test is assessing students' intellectual skills. A second variation on language testing in the Dutch context outlined by Hulstijn (2011) is that adult non-native speakers of Dutch can choose between two state exams which test their language abilities. One is intended to be taken by candidates with lower levels of educational attainment while the second one has been developed for adults with higher levels of education.

2.5.4 Possible changes to standard setting: Weak versus strong performance tests

Who should evaluate test takers from an ethical point of view? Knoch and Macqueen (2016) identified four issues in assessing languages for professional and occupational purposes, namely authenticity, context-sensitivity (for the profession), background knowledge (industry specialists' assessment of language versus language specialist ratings) and performance rating (what is valued by industry), namely who is best placed to judge a performance, domain experts or language-trained raters (Chalhoub-Deville and Deville, 2005; Knoch & Macqueen, 2016; McNamara, 1996, Pilcher & Richards, 2017). Weak performance tests have tasks which reflect the target language use (TLU) domain, but they are assessed by language experts who use linguistic criteria. The problem with using linguistic criteria is that test takers are measured against the 'expert' (maybe non-existent) language user. Strong performance tests include real world tasks and are judged by real world criteria. In these types of test success is guaranteed by task performance with adequate linguistic performance (Knoch & Macqueen, 2016). A strong approach, therefore, takes into consideration the views of professionals in the target language domain, who focus more on communication and the delivery of content rather than linguistic accuracy and complexity. Pilcher and Richards (2017) challenge the power of testing systems such as IELTS because of this focus on language proficiency as opposed to communication, arguing that test scores do not equate to university preparedness and that using the views of subject lecturers is more appropriate. This is because subject lecturers value communication in the context of language use.

On the other hand, the importance of using language trained native speaker raters as opposed to non-language trained native speaker raters is highlighted by Ryan (2016) in the context of asylum seekers to EU countries. This is because the difference in approaches taken by non-experts can lead to wide variations in judgements about language abilities. In addition, the inequity of exempting native speaker test takers from sitting a language test was explained by Knoch and Macqueen (2016) who researched language proficiency requirements for pilots and air traffic controllers. The International Civil Aviation Authority (ICAO) introduced a language policy without clear guidelines with the result that language testing was conducted in an unsystematic way across the authority's 190 member states. A number of tests were developed for the aviation context since testing in this area was deemed to be lucrative. Crucially, Knoch and Macqueen (2016) argued that exempting native speakers meant that this language testing was not conducted ethically, especially when breakdowns in communication were often caused by native speakers not using the correct codes and simple English in emergency situations.

While a combination of language trained raters and domain experts may make the best judgements regarding language competence for the workplace, as in the study by O'Neill et al. (2007) into acceptable language for the nursing profession, this may not be practicable. A more realistic solution may be to disseminate language assessment literacy amongst test users so that they are aware of the, sometimes unattainable, nature of language proficiency. Another solution, not discussed in the research literature, may be to change the marking rubrics in tests such as IELTS so that there is a stronger focus on communication, rather than linguistic accuracy and complexity.

2.6 Changes needed in language assessment literacy

2.6.1 The role of test users

The issue for test users (decision makers) is how to interpret the test results and how to be aware of the test's limitations. Lanteigne (2006) advocates gaining information from test users as regards the type of real-world tasks that should be used. To make ethical decisions, Brown (2014) points to the need for tests to mirror the global reasons for learning English, for test-developers to clearly explain what the test measures and to delineate the sorts of decisions that test users can fairly and legitimately make when confronted with test takers who have received different band scores. In an ideal world, test users would demand change, but these people generally lack the expertise and time to react as professionally as they would like with regards to test results. It may, therefore, be incumbent on test developers (Brown, 2014) to alert decision makers (test users) to the possibility of misuse of test results.

2.6.2 The role of test developers

Test developers are not always concerned with how the test is used by test takers and test users, but they need to be vigilant of the information they disseminate regarding the test's limitations, especially as test taker diversity and test use increases (Chalhoub-Deville, 2016; Kunnan, 2018; Ryan 2016). This is especially the case when test developers are marketing the test for new purposes (Murray et al., 2014). Caines et al. (2014) ultimately call for more communication between the three aforementioned groups. The authors set a challenge for test developers: to develop "easily interpretable materials" (Caines et al., 2014, p. 16) for gatekeepers to use appropriately, rather than making judgements based on "blind faith" (Caines et al., 2014, p. 14). Likewise, Brown (2014) claims it is important not just to consult stakeholders but also to "take them seriously to get them to buy into the curriculum and testing" (p.20). This position was reiterated by Lo Bianco (2018) who advocated the need for

language testers, who have knowledge, to help policy makers, who have power. Although not specifically mentioned by Kane (2013), greater language assessment literacy is increasingly called for when authorities use language tests as instruments of power and control (Ryan, 2016).

2.7 Developments in validation research

In the last twenty years language testing researchers have noted how the terms validity and validation has developed in relation to testing systems. Therefore, in this section, the evolution of the conceptualisation of validity and the ensuing use of validation arguments will be described.

2.7.1 Validity terms are unified

According to Chapelle (1999), validity has been used to indicate the quality or acceptability of a test in education, medicine and psychology. However, the term 'validity' with regard to assessment has engendered much debate. At the beginning of this research it seemed to me that there were a myriad of terms relating to validity: face validity, content validity, construct validity, predictive validity and concurrent validity. Out of these terms, the American *Standards for Educational and Psychological Testing* or the *Standards* (1999), recognised content, criterion related and construct validities. These types of validity focused primarily on the psychometric analysis of individual traits. In 1974 Messick unified validity by amalgamating the three terms (content, criterion and construct validities) into one (validity as a unitary umbrella term comprising all previous types of validity), and he added the areas of interpretation of test scores and social consequences of tests (Kunnan, 2018). The *Standards* were up-dated in 1999 to take into account Messick's (1989) ideas on validation in which he stated that "validation is a process of inquiry in which evidence is marshaled and arguments are put forward for or against in terms of the uses of assessment scores" (Kunnan, 2018, p. 39). For this reason, Riazi (2018) has claimed that validation research can be divided into two eras: pre Messick and post Messick, with validity being unitary, albeit multifaceted. The most recent approaches used to assess educational or psychological tests tend to look at validation as a process within a social arena as opposed to validity which focuses on the internal quality of a test and a test taker.

2.7.2 The argument-based approach to educational test validation

When I was first reading literature on test validity, the terms validity and validation seemed to be used interchangeably. However, Johnson (2011), Kunnan (2018) and McNamara and Roever (2006) point out that validation is linked to the idea of argument or debate. In practical terms, this means using claims or inferences for each aspect of a testing system which needs evaluation. The type of inferences needed will depend on whether the test is being developed or whether it has been used for a long time, the context in which it is being used and who is using the test results. In an argument-based test validation framework, claims or inferences are supported by a warrant. According to Chapelle (2010), this warrant represents an underlying truth or principle, which can help the researcher establish the quality of his/her argument. Each warrant rests on a list of assumptions or statements used in an argument to gather evidence. The research data are then used to support or rebut the assumptions under each warrant (Bachman & Palmer, 2010; Chapelle, 2010; Chapelle et al., 2008; Johnson & Riazi, 2017; Knoch & Chapelle, 2018).

For educational assessments, the argument-based approach is an evaluation of a proposed interpretation and the use of test scores, the reasoning behind the plausibility of this interpretation and an analysis of the consequences (Riazi, 2018). Validity arguments are not necessarily new as they were mentioned by Cronbach and Meehl (1955) and by Messick (1989). Indeed, the term 'validity argument' and 'validation' were both used by Cronbach (1988) when he wrote:

> I invite you to think of 'validity argument' rather than 'validity research'… Validation speaks to a diverse and potentially critical audience; therefore, the argument must link concepts, evidence, social consequences, and values … Questions about tests originate in five perspectives: the functional, the political, the operationist, the economic, and the explanatory. Validators should take seriously questions of all these types" (pp. 4-5).

According to Kunnan (2018), the argument-based approach to evaluating educational assessments has come to the fore in the last twenty years. Taking the above quote by Cronbach (1988), the key words that appertain to my research are the political and economic ramifications of test score interpretations and use, in other words the societal impact of high stakes language testing. Research into the social, economic and political impacts of language

testing is crucial since tests such as IELTS are being promoted as being essential for business, professional registration and migration purposes.

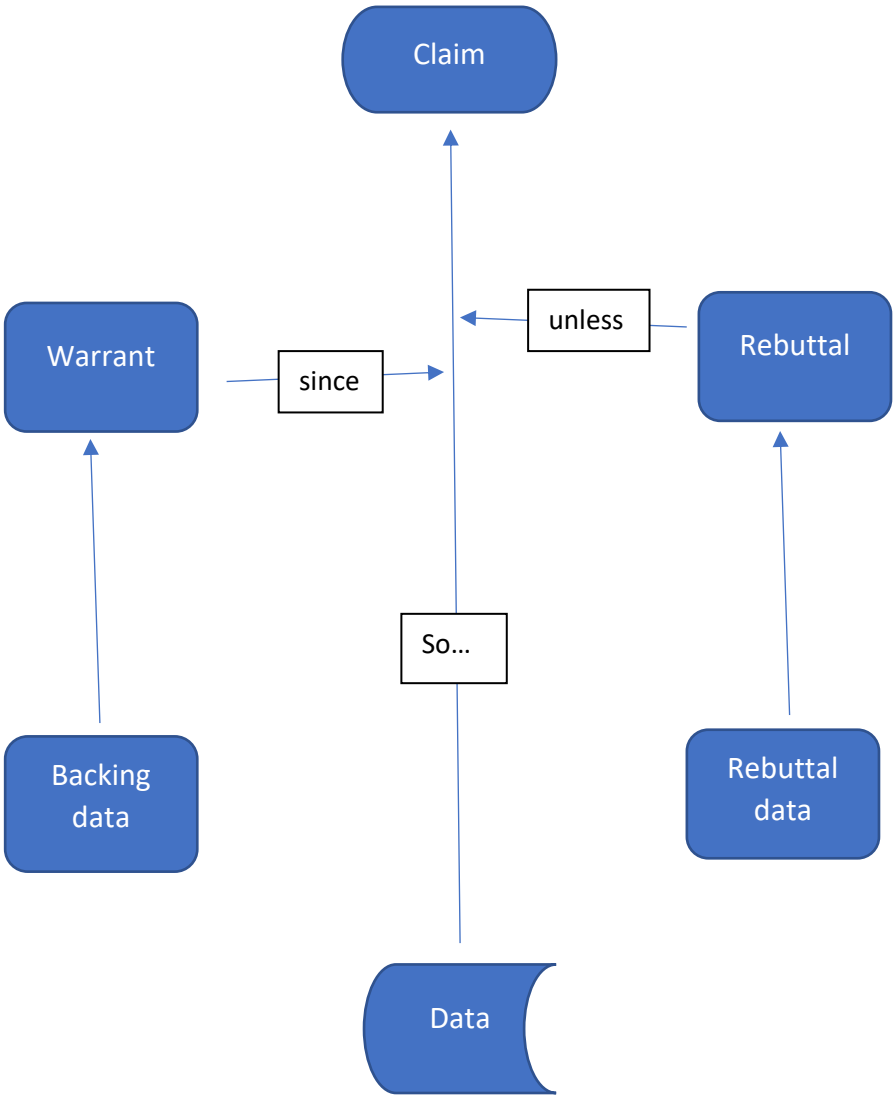
The main reasoning used in an argument-based approach is Toulmin's (1958) philosophical argumentation model which employs syllogisms to create a means to analyse claims using the above-mentioned grounds, warrants, backing and/or rebuttals. A validity argument is, therefore, an interpretive argument accompanied by reasoning in the form of evidence while syllogisms are models of deductive reasoning which contain major, middle and minor terms and generally move from the general to the particular. An example given by Chapelle (2010) includes the inference or claim that a student's speaking abilities are inadequate for studying in English (a general or major inference). The warrant (as in Figure 1) or underlying principle is that frequent hesitations and mispronunciations define students with low levels of English (middle terms). The evidence (minor or particular terms) is collected when students give a presentation which includes a large amount of hesitations and mispronunciations. This evidence may be supported by reflections from a teacher or assessor (the backing data in Figure 1). The conclusion can, however, be rebutted if the presentation topic is highly complex and the student has no knowledge of the topic (rebuttal data, Figure 1). This may be true for hesitations but Chapelle (2010) does not elaborate on why this would be the case for mispronunciations or unnatural pronunciation at the suprasegmental level. Nonetheless, this example demonstrates how the argumentation model uses logical reasoning. In fact, an argument-based approach to test validation is similar to argumentation in a court case (Cook et al., 2015; Davies & Elder, 2005, McNamara & Roever, 2006). In a law suit, each lawyer decides on a general principle to present to the jury and then chooses their warrants and assumptions based on the case at hand. They may choose warrants and assumptions that have been used in previous cases (precedents) and adapt them. Backing evidence presented to the jury lends either support or leads to a rebuttal of the warrant.

Rather than relying on the test construct or a psychometric evaluation of a test taker's traits of abilities, the validation argument looks at the testing process as a social phenomenon. Cook et al. (2015) explained that performance in a test is important, but it is more important to measure or anticipate likely real-world performance. The validation process entails stating the claims or inferences that can be made about a test, based on the context in which the test is being used and the purposes for which it is being used. There are many inferences, warrants and assumptions that may be chosen for any argument-based test validation study. Cook at al. (2015) recommend selecting those elements which are most salient to the intended uses or decisions being made about a test. The inferences are then evaluated using the validity

argument framework in a process which is similar to stating and testing a hypothesis (Chapelle, 2010).

The inferences or claims can relate to test development or test use although test development often relates to the constructs being assessed, test administration and scoring procedures (Kane, 2013) as opposed to the wider social implications of test score interpretation and use, as lamented by Chalhoub-Deville, 2016). Since most high-stakes language tests are already in use arguments relating to consequences rather than test development are more applicable for my study. Furthermore, Cook et al. (2015) claimed that the implications and associated decisions are the most important inferences in the validity argument, which means that analysing the social, economic and political consequences should be a vital element of any argument-based validation research.

Figure 1: An overview of Toulmin's Argument Model (adapted from Bachman, 2005, p.9)



2.7.3 Strengths of Kane's argument-based approach (IUA)

Kane (2013) discussed the implications of using an argument-based approach to validate the interpretation or use of test scores in education in general (classroom-based literacy, maths and science tests). However, he makes a number of references to high-stakes exams and the wider implications of the results. As Kane (2013) alleges "Concerns about validity have their roots in public concerns about the appropriateness, or fitness, of the test scores for their intended use or uses" (p.2). Kane (2013) expands on Bachmann and Palmer's (2010) framework which concentrated on using test score uses to assess test validity by adding **interpretations** of test scores. He labels his framework as an interpretation/use argument (IUA), stating that if test scores are going to be used to predict future competence or success in an employment or educational context, there needs to be evidence that any predictions are accurate. For Kane (2013), "Validity is not a property of the test. Rather, it is a property of the proposed interpretations and uses of the test scores" (p.3). The process he advocates is, on the face of it, simple: the first step involves stating the claims being made by looking at the links between test performance/test scores and the decisions being made about them (how they are commonly interpreted and used by decision makers) in enough detail to build a framework. The second stage entails assessing the plausibility of these decisions in relation to what the test scores actually mean.

2.7.4 Limitations of Kane's argument-based approach

Kane's (2013) arguments are developed logically and in-depth. He states that the argument-based approach is simple because the only requirements are that research looks at the claims and predictions being made and then evaluates the plausibility of those claims/predictions. However, he does not give a practical example, so the depth of investigation needed is unclear, while the level of detail made the approach complicated and not easy to implement in a practical sense. Kane (2013) does state that an IUA needs judgement, partly to avoid taking some assumptions in an argument for granted, which leads to an under-analysis or the opposite, trying to overstate the problem by attacking testing systems. He points out that "it can be tempting for hostile critics to overstate assumptions inherent in the testing program's claims (e.g. by asking licensure programs to provide predictive validity evidence); more-ambitious IUAs, like "straw men", are easier to knock down than less-ambitious IUAs" (Kane, 2013, p.18). One example of a straw man fallacy Kane (2013) cites is asking licensure tests to predict performance in future practice. He says that this could be useful for employment decisions but that it is difficult to validate in practice. Yet, it seems that this is what test users

expect from test scores; they make extrapolation inferences about a score and possible 'real-world' performances (Merrifield, 2008; 2011). In reality, on the job competence requires a variety of 'observable attributes' or traits and skills, of which language proficiency is only one.

Should the consequences of score uses be assessed in terms of validity or utility? Kane (2013) firmly believes that test score uses and interpretations are related to validity and that it is necessary to have a clear statement of the proposed interpretations and uses, followed by a critical evaluation of them. However, Kane (2013) states that technical skill and ingenuity are needed to create an IUA, a skill which may be beyond many researchers. This has made the use of an IUA difficult to implement. Additionally, Kane (2013) believes that tests which result in high-stakes decisions should have an extended research programme for validation, an issue which the IELTS partners are addressing.

2.7.5 The use of the argument-based approach for language test validation

Kane's Interpretation Use Argument (IUA) approach to test score interpretation validity was investigated in-depth by Chapelle, Enright & Jamieson (2010) who used the IUA framework to evaluate the validity of the iBT TOEFL test. They welcomed this framework as a means to evaluate language tests. Essentially, this is because the traditional approach to test validity entails assessing the validity of a construct such as communicative language competence, a construct which is difficult to assess effectively due to the complex nature of language proficiency (Chapelle et al., 2010; Rose & Syrbe, 2018; Ryan, 2016). Conveying the parameters of such a construct to test users is even more problematic. As Chapelle (2011) points out "In language assessment, such alignments with frameworks are controversial because they attempt to connect social and political meanings associated with frameworks with the scientific procedures used to understand score meaning" (p. 25). For this reason, an argument-based framework has been increasingly used (Bachman, 2005; Bachman & Palmer, 2010; Chapelle et al., 2008; Chapelle et al., 2010; Johnson & Riazi, 2017; Knoch & Chapelle, 2018). Kane's (2013) IUA was complicated in practice so these authors have made efforts to operationalise a validity argument framework using inferences, assumptions, warrants and rebuttals for the context of a variety of language tests.

A validity argument framework can include inferences or claims such as evaluation, generalisation, explanation, extrapolation, decision and consequence inferences (Knoch & Chapelle, 2018). Other inferences could be domain definition, utilisation, ramification, authenticity and positive impact (Chapelle, Cotos & Lee, 2015). Johnson and Riazi (2017) demonstrated how an argument-based framework can be used in a feasible way and how Kane's interpretative model which used

evaluation, generalisability and extrapolation inferences could be combined with Bachman's and Bachman and Palmer's assessment use argument (AUA) by adding inferences related to decisions and consequences. Although Johnson and Riazi (2017) encountered problems by including numerous stakeholders in their research, they stated that this enhanced the 'ecological-validity' (p.102) of the study by including a variety of perspectives.

However, the terminology used by such a variety of authors in not consistent or well defined, maybe because an argument-based approach to language testing is a nascent area. For example, the words 'inference' and 'claim' are used differently by researchers. Bachman and Palmer (2010) prefer to use the term 'meaningfulness' instead of validity, claiming that it is a less technical term and should therefore be more comprehensible to the wider public. Evidence that can be gathered to determine language assessment meaningfulness includes evidence that a test can predict future real-world performance in a similar domain (criterion related or predictive evidence) and evidence that a test has meaningful consequences for the community (Kunnan, 2018). The hope in using test meaningfulness (or validity) is that validation studies can act as road maps to improve stakeholders' decisions made on testing (Johnson & Riazi, 2017).

Another drawback in utilising an argument-based approach to test validity is that researchers lay responsibility on different stakeholders to make testing systems more equitable. For example, Johnson and Riazi (2017) state that "test users in particular are often if not typically negligent in investigating both the meaning of test outcomes and the impact of test use for their particular educational context" (p.85) whereas other researchers (Chalhoub-Deville, 2016; Kane, 2013) claim that it is incumbent on test developers to ensure tests are used ethically. The opportunity, however, of taking an argument-based approach is that inferences used by different researchers can be omitted, combined and new ones created for specific research contexts.

2.7.6 How to operationalise an IUA for immigration, employment and professional registration in the New Zealand context

As Kane (2013) and Kunnan (2018) point out, people who use test scores are not always those for whom the test was originally developed "interpretations and uses can change over time in response to new needs and new understandings" (Kane, 2013, p.37). Knoch and Macqueen (2016) claimed that when a test is used for a different purpose to which it was originally designed, the interpretations of the scores may not be valid. In the case of IELTS, the partners (IDP Australia, Cambridge ESOL and the British Council) who developed the test are currently marketing it for business and migration purposes rather than academic study alone (Merrifield, 2008; 2011). Scores

are then used by test users, such as employers, to make decisions about a person's employability. When this happens, test users are making '*generalisation inferences'* from a test score to "claims about expected performance in a universe of possible observation" (Kane, 2013, p.10). As Kane (2013) points out, such inferences are usually presumptive so require support. Therefore, Kane (2013) alleges that "it is important to specify in some detail those parts of the IUA that are central to the proposed interpretation and use and that are in serious doubt" (p.15) to ensure that score-based decisions do not lead to negative consequences.

As Caines et al. (2014) conclude, concepts of validity for a testing mechanism are evolving as the world is also being transformed into a global marketplace. Kane (2013) points out that the validity of test score interpretation and use is strengthened if it is reliably interpreted by stakeholders. According to Caines et al. (2014) and Kane (2013) test developers and psychometricians should consider the following questions:

- Do the scores achieved reflect the test-taker's present abilities or how they are likely to perform in the future?
- Do the scores reflect how the test-taker can perform in a real-world environment?
- Which scores reflect the fact that a test-taker is minimally competent?
- How much variation will gatekeepers tolerate?

This is especially important in contexts where "legislators and policy makers intend to use assessments as instruments to further public policy" (Kunnan, 2018, p. 209). Caines et al. (2014) claim that test users take test data or test scores as a basis for making a chain of inferences which leads to a decision rule. In line with Toulmin's (1958) approach, such decision rules should be analysed in terms of how appropriate all the inferences in the chain and the ensuing decision rules are. According to Kane (2013), the decision is warranted if the rule works well for the target population and any unintended consequences are avoided. Such consequences can be viewed negatively if they have "substantial effects on a significant number of individuals" (Kane, 2013, p.48) and in this case would lead to a rebuttal of the validity inference or claim.

It would be appropriate, therefore, to ascertain if decision rules are made in New Zealand with the use of accurate information regarding the constructs measured in an IELTS test. Test developers do not make the rules, but they need to be seen as acting responsibly (Caines et al., 2014) in contexts which change in line with social and political agendas. Hulstijn (2011, 2015) emphasises that test developers and the testing industry in general, should be

transparent regarding the intellectual skills that are needed to attain the highest levels in what is marketed as a language test. Ultimately, any negative consequences of a decision rule have to be weighed against the positive consequences of using a testing system, which leads to a trade-off. Resolving this trade-off can be difficult in practice but is particularly crucial if high stakes decisions are being made (Kane, 2013). Kane contends that a strong validity argument does not necessarily justify a decision rule; rather the question of "whether the rule makes sense and is justified in a particular context for a particular purpose" (Kane, 2013, p.56) should be addressed.

Kane (2013) advises that if the test is being used to determine who is suitable for a job, it would be appropriate to start by delineating some of the skills and aptitudes that are necessary in that field of work and then "develop a test that seems to measure these attributes" (p.16). Test design should theoretically be influenced by the proposed interpretations and uses of the test. However, IELTS does the reverse by using a test which was developed for academic purposes and then marketing it for business and migration. Both Kane (2013) and Kunnan (2018) advise against making claims about test validity after the assessment has been developed (post hoc). Yet, if the test is being used for new purposes it is necessary to evaluate its validity for these ends.

2.8 The societal impacts of language tests

2.8.1 The fairness and justice argument

Kunnan (2018) hoped to expand the argument-based approach to the validity of tests by including ethical reasoning concerned with fairness and justice, namely he introduced a framework to analyse how fair tests are and how just the institutions are that develop the test. He looked at a wide variety of educational tests in different cultures (Australia, China, the UK and USA) and concluded that even though test developers may have the intention of producing fair testing systems, in reality they often do not. Some of the tests he probed include civil service exams used for hundreds of years, classroom tests (including No Child Left Behind and Race for the Top), as well as immigration and citizenship tests. Kunnan's focus on the need for testing institutions to be just seems rather utopian, although some of the examples he gives reflect a dystopian use of educational tests in the sense that they are used in a discriminatory way and to mask hidden, usually political, agendas (Kunnan, 2018). Also, the emphasis on fairness can be misleading as this term may not have the same meaning in a variety of social, cultural, economic and political arenas. The two areas that are pertinent for

my research are the societal impacts of language tests, such as the exclusion of competent employees for the workforce, and Sen's (2009) argument that outside perspectives are needed, together with public reasoning to advance fair assessment practices.

Kunnan (2018) suggested that tests are often not challenged, mainly because people believe they are created by 'experts' and, for this reason, are seen as being infallible. He contends that test developers may well try to avoid societal consequences and may see this as outside of their remit but, as Cronbach (1988), pointed out people concerned with test validation should be obliged to reflect on test consequences. The societal implications of tests are now included in the latest AERA, APA, and NCME *Standards* (2014) and the International Language Testing Association, ILTA (2018) guidelines. The *Standards* mention the importance of accurately informing policy makers about test characteristics while ILTA addresses the societal roles of language testers, their obligations to the society within which they work and the potential effects of language testing on all stakeholders.

According to Sen (2009), the role of outsiders and public reasoning can prove to be beneficial in the debate about just institutions because they are more impartial than insiders. The role of language assessments may often be debated by language assessment professionals but there is very little public reasoning about the meaning of language tests, how useful they are and the impacts on society. However, as Kunnan (2018) claims "an assessment institution ought to be just and bring about benefits in society and promote positive values and advance justice through public reasoning" (p.94). The two groups that may be seen as contributing to public reasoning about language testing for migration to New Zealand from my research are the native speakers of English who took a test of English for research purposes and the New Zealand employers. Both groups of research participants hold outside perspectives about language assessment for migration and can arguably add practical elements to the debate.

While Kunnan (2018) concentrates predominantly on the fairness of tests in the design and development stages as opposed to the intended and unintended consequences for societies, he does state that fairness and justice needs to be addressed by all stakeholders, including test takers, who:

> are the centrepiece of the assessment enterprise and hopefully the beneficiaries of assessments. Finally, there is the group that is arguably the most powerful of the lot: institutional administrators, government officials and policy makers (p. 225).

For this last group Chalhoub-Deville (2016), Kunnan (2018), and Lo Bianco (2018) suggest the need for on the job training. According to Sen (2009), public reasoning of assessment validity can help elucidate if an assessment is beneficial for the immediate stakeholders, wider stakeholders and the whole community or nation. An important point that Kunnan (2018) notes is that "assessments for immigration and citizenship are not accountable to immigrants (as they have very few rights in these matters" (p.211).

2.8.2 Theory of Action / Social consequences of tests used for immigration, employment and professional registration

Taking the issue of fairness in testing one step further, researchers have more recently been concerned with the social impact of language tests being used for immigration and citizenship and the connection with policies for would-be migrants (Chalhoub-Deville, 2016; McNamara & Ryan, 2011; Ryan, 2016; Shohamy, 2001; Shohamy & McNamara, 2009). Ryan (2016) advocates future research be focused on the *reasonableness* of testing systems for citizenship and the effects of these tests on democracy at the individual and societal levels. Shohamy (2001) argues for a Critical Language Testing (CLT) approach to language testing research which investigates the uses and misuses of language tests and endeavours to empower test takers. Chalhoub-Deville (2016) posits that in today's world of global educational reforms it is necessary to have research into the interconnections between policy mandates, testing systems and societal consequences. The policies and underlying ideologies in Ryan's (2016) words "present the greatest challenge to researchers, not necessarily because they are particularly difficult to expose, more because they are particularly difficult to shift" (p. 321).

The consequences of tests and how they are used have been mentioned as factors in evaluating test validity by Messick (1989) and Kane (2013) but their analyses have been more retrospective, technical and psychometrically based according to Chalhoub-Deville (2016). This can be explained by the early emphasis on how test constructs measure human traits from a psychological point of view with a focus on the individual rather than the social consequences. Messick (1989) allows for adverse consequences but expects these to be related to parts of the test which include construct misrepresentation (either construct underrepresentation or construct irrelevance). While Kane (2013) acknowledges that adverse consequences affect test validity, he focuses more on test score interpretation and use at the testing level rather than the wider social impact. For Chalhoub-Deville (2016), Kane's (2013) IUA and Messick's (1996) works have been significant but lack a social perspective.

To address the lack of social impact, Chalhoub-Deville (2016) outlines the need to research the socio-political consequences of validity arguments by looking at the consequences of score interpretations and uses at the individual level, at the aggregate level and the systems level (the educational or social contexts). This is meant to lead to a proactive approach to testing validation and policy research by discussing social impact assessment (SIA). The roles and responsibilities of individual, aggregate and educational-social groups are analysed with the help of a *zone of negotiated responsibility* (ZNR) in the hope of discovering if a testing system is reaching its intended educational-economic-social goals. One point, that has already been frequently mentioned, is that the use of tests may change over time. What often evolves is that test users create new interpretations of/uses for results. Test-creators cannot ignore the fact that changes in the interpretations/uses have become common practice, especially as they are likely to profit from the increased number of tests being administered (Chalhoub-Deville, 2016). This is what has happened with language testing in New Zealand as there used to be two or three testing centres, but this has now proliferated to around twenty centres.

2.8.3 Linking test validation research with policy and policy makers

Chalhoub-Deville (2016), McNamara (200), McNamara and Roever (2006) claim that researchers into language assessment have neglected 1) policy research, 2) language testing as a social and political practice, 3) the social context of testing and 4) that very few Theory of Action (TOA) articles have emerged, so that provides a good reason for undertaking such research in New Zealand. According to Chalhoub-Deville (2016), "consequential research at the policy development stage is practically non-existent …. Publications in the area do not accommodate the changing functions of assessments and the validity research responsibilities they require" (p. 468). Apart from the development stage, that was regarded as pivotal by earlier validity researchers, Chalhoub-Deville (2016) states that research should continue once assessment is live. Research into high stakes language testing systems such as IELTS and TOEFL are perforce reactive as the tests have been administered since the late 1980s but what is new is that IELTS is being promoted as being the test for business and migration purposes.

It is increasingly becoming important for researchers in language testing to be involved in the dissemination of increased language assessment literacy for, amongst others, classroom teachers and policy makers (Lo Bianco, 2018). Elder (2018) was part of a successful opposition to the Australian government who wanted to introduce a requirement of IELTS 6.0 to be able to sit the citizenship test. This policy has now been revoked. Deygers (2018) has questioned if it is justifiable to

use language testing to restrict another's freedom, capabilities and opportunities, especially when there are 244 million migrants worldwide.

Any policies which are linked to language assessment should, therefore, come under scrutiny. Policy analysis lends itself to interpretive approaches such as hermeneutics (Diem & Young, 2015; Levinson, Sutton & Winstead, 2009) because policies are socially constructed and the subject of the policy (language assessment in my case) is also a socially constructed phenomenon. As a result, any solutions to unintended/overlooked consequences need to be socially constructed. Although Sen (2009) promoted incorporating a global perspective, policies are created by communities of practice (Lave & Wenger, 1991) and are appropriated by local actors who add their own sense making of what the policy means and how it should be implemented. As Levinson et al. (2009) point out, IELTS has been reified and taken on a life of its own in New Zealand policy making but the test and therefore the policies may not be understood by the community of practice who use them.

2.9 The need for research into language tests for workplace communication

In summary, the use of high-stakes language tests such as IELTS has expanded greatly in the last decade with an ever-increasing number of test takers, test centres and test users accepting score results. While IELTS was initially designed for entrance to tertiary study, it is now being marketed as a passport to succeed in the workplace in English speaking countries. Therefore, the validity of the test for the workplace needs to be evaluated.

Another issue is that of responsibility for decision rules. While the test users may be in the best position to assess their decisions rules, the test developer has some responsibility to help test users interpret the test, especially if they market the test for specific purposes by claiming it measures certain attributes. In this case, the onus may be on test developers to justify the test validity in new circumstances (Murray et al., 2014).

IELTS test scores reflect a person's English language proficiency, and this is assessed by referring to discrete item lexical/grammatical accuracy as compared with an 'expert' native speaker standard. The differences between BICS (basic interpersonal communication skills) and CALP (cognitive academic language proficiency) as defined by Cummins (1984), or BLC (basic language cognition) versus HLC (higher language cognition) as described by Hulstijn (2011), are not taken into account.

For ease of interpretation, there may need to be more consultation with test users (language assessment literacy education), and the provision of cogent, easily understood materials.

Murray et al. (2014) place the onus on the test creators when they allege that "if a test is being used for a purpose other than that for which it was initially conceived and developed, the test provider bears some ethical responsibility for the investigation of these effects" (p.65). For O'Loughlin the key research question was: "To what extent are IELTS test scores used in valid and ethical ways for the purpose of university selection?" (O'Loughlin, 2008, p 150). For my purposes, the question which is relevant is: To what extent are language test scores used in valid and ethical ways for the purpose of immigration, employment and professional registration?

Aryadoust (2013) used a validity argument to analyse the claim that IELTS ensures the right level of English to succeed in study, work or social spheres using evaluation, explanation, generalisation and extrapolation inferences and concluded that the listening module does not adequately assess listening constructs for academic study or vocational purposes. Aryadoust (2013) theorised that the listening module cannot be regarded as being valid for all the areas in which the test scores are used. The perceptions of native speakers who work in New Zealand regarding test content in all four skills can add to the debate about whether test scores can reliably be used to identify test-takers' performance in non-academic contexts, especially as these contexts have been under researched.

Therefore, in this research I have endeavoured to widen the debate about using a test such as IELTS for non-academic purposes to shed light on one of the key claims made by the test developers that "IELTS ensures the right level of English to succeed in study, training or life in a new country" (ielts, 2019). If all the above inferences are supported by the participants in this research, language assessment results are being interpreted and used in New Zealand in a valid and ethical way.

2.10 Overview of previous research into the use of high-stakes language tests

Table 1. Some of the previous research into the use of high-stakes language tests

Authors and publication	Title	Focus
Al-Issa, A., Al-Bulushi, A. H., & Al-Zadjali, R. (2016). *The Qualitative Report*, *21*(5), 848-863.	Arab English language teaching candidates climbing the IELTS mountain: A qualitatively driven hermeneutic phenomenology study.	How are the learning strategies of prospective English language teachers taking IELTS in Oman affected by their learning ideologies? These ESOL teacher trainees struggled to gain an IELTS score of 6.5.
Brown, J.D. (2014).	The future of world Englishes in language testing.	With increased mobility, English language testing systems take the

Language Assessment Quarterly, 11(1), 5-26.		existence of world Englishes into account in a global environment?
Caines, J., Bridglall, B. L., & Chatterji, M. (2014). *Quality Assurance in Education*, *22*(1), 5-18.	Understanding validity and fairness issues in high-stakes individual testing situations.	Do all stakeholders of a testing mechanism perceive the issues of validity and fairness in the same way? The three stakeholder groups discussed were: test makers, test takers and decision makers.
Elder, C. (1993) *Melbourne Papers in Language Testing, 10*, 235-254.	Language proficiency as a predictor of performance in teacher education.	IELTS test score results may be relied on to indicate performance on teacher education courses in the short term, but not in the long term. A score above 4.5 with 5.5 in the listening section may be adequate.
Gribble, C., Blackmore, J., Morrissey, A-M., & Capic, T. (2016). *IELTS Research Reports Online Series, 2016-2*.	Investigating the use of IELTS in determining employment, migration and professional registration outcomes in healthcare and early childhood education in Australia.	This research focused on overseas-trained professionals and international graduates transitioning into the health and early childhood education sectors in Australia. The authors canvassed the perceptions of employers, employees, government and professional bodies in Australia.
Merrifield, G. (2008). *IELTS Research Reports, 8.*	An impact study into the use of IELTS as an entry criterion for professional associations – Australia, New Zealand and the U.S.A.	This research sought to ascertain why professional associations chose to use IELTS as a testing mechanism of language competence for entry into employment and to compare it with other language assessment systems. A third area analysed was the perceptions of these associations regarding the level of support provided by the IELTS partners.
Merrifield, G. (2011). *IELTS Research Reports, 11.*	An impact study into the use of IELTS by professional associations and registration entities: Canada, the U.K. and Ireland.	The objective of this research was to investigate the reasons why professional bodies opted to use IELTS as an entry criterion, the

		minimum scores they deemed to be acceptable and to outline the main competitors to IELTS in Canada, the U.K. and Ireland.
Moore, T., Morton, J., & Price, S. (2011). *IELTS Research Reports, 11.*	Construct validity in the IELTS academic reading test: A comparison of reading requirements in IELTS test items and in university study.	The authors sought to compare the construct of reading tasks at tertiary level with the skills and abilities tested by IELTS. The perspectives of academic staff at Australian universities were analysed.
Moore, T., Morton, J., Hall. D., & Wallis, C. (2015). *IELTS Research Reports Online Series, 2015-1*.	Literacy practices in the professional workplace: Implications for the IELTS reading and writing tests	The aim of this research was to ascertain if IELTS reading and writing tasks predict readiness for professional employment by analysing the constructs tested by IELTS and comparing them with the constructs needed in a wide range of professions. The authors researched the perspectives of employers and supervisors in the following workplace domains in Australia: accounting, education, engineering, IT, law, management/administration, media, health and science.
Murray, J., Cross, J.L., & Cruickshank, K. (2014). *IELTS Research Reports Online Series, 2014-1.*	Stakeholder perceptions of IELTS as a gateway to the professional workplace: The case of employers of overseas trained teachers.	This research explored the perceptions of school principals regarding the IELTS test as a gateway to employment for overseas trained teachers. A major conclusion was that stakeholders need to be better informed of what English proficiency tests can and cannot assess.
O'Loughlin, K. (2011). *Language Assessment Quarterly,8*(2), 146-160.	The interpretation and use of proficiency test scores in university selection: How valid and ethical are they?	This research analysed how test users (administrative and academic staff) at tertiary institutions in Australia interpreted and used IELTS test score results. The author found

		serious flaws in the interpretation and use of test scores for university entrance.
O'Neill, R., Buckendahl, C.W., Plake, B.S., & Taylor, L. (2007). *Language Assessment Quarterly*, *4*(4), 295-317.	Recommending a nursing-specific passing standard for the IELTS examination.	Setting a legally defensible standard of English language proficiency for overseas-trained nurses in the U.S. using IELTS. The panel of experts included ten nurses who had English as a second language and had themselves taken IELTS, seven U.S. educated nurses who worked with patients who spoke a second language, together with nurse educators, supervisors and regulators with one panellist being a member of the public.
Pilcher, N. & Richards, K. (2017). *Power and Education, 9*(1), 3-17.	Challenging the power invested in the International English Language Testing System (IELTS): Why determining 'English' preparedness needs to be undertaken within the subject context.	The perceptions of lecturers regarding the 'English' of IELTS compared with the 'English' needed at tertiary level in the I.K. in the areas of design, nursing, engineering, business, computing and psychology. The authors argue that students' English readiness should be determined by subject specialists rather than IELTS raters.
Read, J., & Wette, R. (2009). *IELTS Research Reports, 10.*	Achieving English proficiency for professional registration: The experience of overseas-qualified health professionals in the New Zealand context.	This study explored the experiences of overseas-trained health professionals who were taking IELTS and the OET to meet the English language requirements for registration in New Zealand.
Roshid, M.M., & Chowdhury, R. (2013). *Mevlana International Journal of Education, 3*(1), 68-81.	English language proficiency and employment: A case study of Bangladeshi graduates in Australian employment market.	What is the relationship between English language proficiency as assessed by IELTS and employability in an Australian context? The perspectives of IELTS test takers.
Rumsey, M., Thiessen, J., Buchan, J., & Daly, J. (2016).	The consequences of English language testing for	The original intention of the authors was to investigate the

International Journal of Nursing Studies, 54, 95-103.	international health professionals and students: An Australian case study.	mobility of people employed in the health sector in Australia but as the research progressed it became apparent that English language testing was seen as a barrier so the research transformed into an analysis of the impact of IELTS on migration and the practices of health professionals in Australia.
Sawyer, W., & Singh, M. (2011). *IELTS Research Reports, 11.*	Learning to play the 'classroom tennis' well: IELTS and international students in teacher education.	What is an appropriate IELTS score for graduate entry teacher education courses? The perspectives of lecturers, teacher registration authorities and students in an Australian context.
Sedgewick, C., Garner, M., & Vicente-Maria, I. (2016). *IELTS Research Reports Online Series, 2016-2.*	Investigating the language needs of international nurses: Insiders' perspectives.	This research investigated the extent to which IELTs scores for internationally qualified nurses are appropriate and adequate for practising as nurses in the U.K. Listening skills, as tested by IELTS, were deemed appropriate but many competencies required to gain a 7.0 in the other three skills were marginally relevant for the workplace while some skills essential for nursing are not tested at all by the IELTS test. The perspectives of international nurses.
Smith, H., & Haslett, S. (2007). *IELTS Research Reports, 7.*	Attitudes of tertiary key decision-makers towards English language tests in Aotearoa New Zealand: Report on the results of a national provider survey	This study found that IELTS is the most frequently used test for entrance to tertiary courses in New Zealand, but that decision makers would like more information about the test and how test score results may be interpreted.
Stricker, L.J. (2004). *Language Testing*, *21*(2), 146-173.	The performance of native speakers of English and ESL speakers on the computer-	A comparison of the scores attained by native and non-native English speakers on the GRE

	based TOEFL and GRE general test	general test and TOEFL. The author concludes that speakers of English as a first language vary in linguistic, education and cognitive skills and do not necessarily outperform non-native speakers of English in English language assessments.

2.11 Contributions of this research

The table presented above (table 1) shows that previous research has focused on the use of IELTS and other high-stakes language tests for the educational and health sectors in Australia, but that little research has analysed the use of language tests for a variety of workplaces in the New Zealand context. Further, the viewpoints of native speaker employers and test takers regarding the type of English that is needed for the modern workplace have been under-researched. Therefore, the first issue of significance for this study was whether IELTS is testing skills that are necessary in the contemporary workplace. Indeed, at the end of her study, Merrifield (2008) recommended that "a process of test validation of IELTS be undertaken, including an analysis of language skills for the workplace for each of the professions represented by the professional associations which use it "(p.29). Although Merrifield (2008) intended to interview candidates who had taken IELTS in their workplace to investigate their perceptions of the relevance of the exam for their workplace interactions, this proved to be impossible at the time of writing but the perspectives of test takers (native and non-native speakers of English) who have employment are crucial to discussions about testing validity. In order to analyse this, it was necessary to assess what types of skills and what types of language test takers and test users regard as being necessary for general workplace tasks.

By using native speakers of English as test candidates and analysing their views about the test has arguably added to the debate since they generally consider themselves to be competent in their mother tongue for work/study purposes. There has not been research that analyses the perceptions of native speakers of English who have taken IELTS, nor which investigates the beliefs of test users (native speaker employers) regarding how they use language test results, which language skills are most needed at work and what they infer about a person's competence for the workplace in relation to their band score results.

According to Hulstijn (2011), the CEFR test developers have conceded that:

> Many adult native speakers will never attain the highest CEFR levels (C1 and C2). However, the authors do not explain why this is so, nor do they explicitly acknowledge that the C1 and C2 levels will generally not be attainable by L2 users with educational backgrounds other than higher education" (p.241).

This is because the native speaker performs ideally in terms of basic language cognition (BLC), not higher language cognition (HLC). It is imperative, therefore, to research native speakers' perceptions of language testing in different contexts. Ultimately, Hulstijn (2011) calls for more transparency from test developers; they should acknowledge that only candidates with relatively high intellectual skills are likely to attain the highest levels on the CEFR (Common European Framework (see appendix R for a CEFR/IELTS score comparison). This is currently the situation with IELTS as well, with native speaker test candidates receiving an average test score result of 6.9 while some test users require non-native test takers to gain a minimum of 7.0 in all four skills. In this respect, an apposite question, as raised by Kunnan (2018), is whether language tests are being used in a fair and just way.

The aspects of Kane's (2013) IUA which are relevant to IELTS are that test scores are a snapshot of a test takers performance under exam conditions on a certain day. Yet, these scores are then used by others to support claims that a test-taker has certain traits and abilities in a particular field and, thus, the likelihood to succeed in an educational or employment situation. According to Kane (2013), to ensure test validity it is necessary to interview test users to see how they make decisions about language test results and how they interpret the information disseminated by IELTS on their web pages as regards the test content, the difference between band scores and how this compares with language which is used in the workplace (TLU).

Validity frameworks have been increasingly used in the area of educational testing since the 1990s (Bachman, 2005; Bachman & Palmer, 2010; Chapelle et al., 2008; Cook et al., 2015; Johnson, 2011; Jun, 2014; Li, 2015). The development of these frameworks has moved the focus from the reliability of scores (a positivist approach) to an analysis of how the scores are interpreted and used by societies, together with the intended or unintended consequences of test use (an interpretive approach). Despite the development of logical and conceptual frameworks to aid research into whether tests are being used meaningfully and ethically,

“many in the literature lament a chronic lack of these essential validation efforts (Johnson, 2011, p.44). This still seems to be the case with a dearth of research which uses argument-based approaches to test validity, and particularly into the societal impact of test use in specific contexts.

In the next chapter (Chapter 3), the rationale for using a qualitative paradigm to study the use of language tests in New Zealand will be presented. This will be complemented by a description of the data gathering and why a hermeneutic approach to data analysis was applicable. The research setting and research participants will be described in depth and issues related to ethics, rigour and trustworthiness in relation to the particular research context will be addressed.

CHAPTER 3

METHODOLOGY AND RESEARCH DESIGN

3.0 Introduction

This chapter describes why an interpretive research paradigm was chosen for this study and how an inductive hermeneutic approach to the analysis of data from semi-structured interviews with test takers and test users helps elucidate the validity and ethics of claims made by the test creators. As pointed out by Snape and Spencer (2003) and Yin (2011), using an inductive or bottom-up approach to data analysis fits the spirit of the qualitative approach. This entails looking for patterns in the data to reflect upon and appreciate the kaleidoscope of perspectives presented by the research participants while not forgetting the researcher's own experience as the researcher's "personal experience is also a source of data about the research problem" (Ezzy, 2002, p. 153). Specifically, the aim of this chapter is to outline how and why decisions were made to take an interpretive qualitative research approach towards test validation to explore the following question: To what extent are language test scores used in valid and ethical ways for the purposes of immigration, employment and professional registration in a New Zealand context?

As with Gribble et al.'s (2016) qualitative research into the use of IELTS by the medical and early childhood sectors in Australia, my purpose was to ascertain the perceptions and beliefs of a variety of test-takers and test-users regarding their knowledge of the IELTS exam and band score parameters, its merits and limitations, and its usefulness for employment and living in the New Zealand context.

Language learning, teaching and testing are not conducted by individuals in isolation. Rather, these practices are collaborative and co-created by multiple players since language learning is essentially a social practice with links to identity (Loewen, 2015; Ortega, 2013; Polio & Williams, 2009). As Ortega (2013) suggests "People who undertake to learn an additional language are engaged in changing their worlds" (p. 250) by gaining linguistic and sociocultural competence, as well as new identities, in the worlds of people who speak a different language. For this reason, interpreting and re-presenting the perspectives of test users and test takers was deemed to be most appropriate for this study. Cook et al. (2015) explained that research into assessment validation has moved into a post-psychometric era of assessment in which qualitative and subjective data are increasingly valued. This is why a qualitative methodological approach and research design guided the collection and analysis of data in this study. The guiding lens behind the analysis was the argument-based approach to test validity in order to show New Zealand stakeholders' perspectives towards test content and test

scores. This argument will demonstrate whether high-stakes language tests are being meaningfully interpreted and ethically used in the New Zealand context. The research design decisions were intended to ensure rigour and trustworthiness. Firstly, the reasons for choosing a qualitative paradigm will be discussed, followed by an explanation of how a hermeneutic approach to thematic analysis was used. The research setting, and participant selection will then be explained, together with the ethical issues related to this study.

3.1 Research approach

The interpretive paradigm fits well into research investigating the experiences of different native and non-native speaker test takers when needing to take high stakes language tests such as IELTS (the phenomenon) for work and/or migration purposes. This is because different cultures can have different ontologies or ways of perceiving what is real in the world and they evaluate the world in an epistemological sense according to certain variables, which are often related to socio, cultural and educational practices. Non-native speaker participants in my research originated from a variety of countries including China, Germany, India, Malaysia, the Philippines, Poland, Singapore and Vietnam. These countries have different educational traditions and socio-economic traditions. For this reason, it was hoped that these non-native speaker test takers, who had taken IELTS in the past and were now working in New Zealand, would have varying beliefs about the value and importance of English language testing. These perspectives, together with those of New Zealand employers, create a kaleidoscope of perspectives, and thus offer "richer and more multifaceted accounts of test properties and score meanings "(Davies & Elder, 2005, p. 810). In addition, stakeholders from different employment sectors and different nationalities may have a variety of ways of assessing the relationships between human beings and society, what counts as legitimate knowledge, and how such knowledge and relationships may be known (epistemological concerns or debates) (Davidson & Tolich, 1999). The qualitative research approach was, therefore, chosen because it allowed me to investigate the experiences of test takers and test users in depth and to provide a rich description of how the participants made sense of high stakes language testing for migration purposes.

3.2 Strengths of an interpretive qualitative approach

The aim of traditional quantitative oriented research was to provide 'facts' in a neutral, objective way. The statistics gleaned from quantitative studies would, it was thought by the majority of researchers, be valid, reliable and generalisable. In fact, depending on the research

question(s), some research is better suited to statistical analysis. However, according to Snape and Spencer (2003), the influence of philosophers such as Immanuel Kant, Wilhelm Dilthey and Max Weber started to permeate research and took a strong footing towards the end of the last century. The assumption that there is only one 'truth' or 'perspective' was seriously challenged, as the aforementioned philosophers claimed that 'truth' is itself a social concept, socially constructed and dynamic.

The first strength of a qualitative research methodology is that it provides a thorough, rich investigation of the many facets of a particular lived experience, such as language testing for migration. Snape and Spencer (2003) claim that a methodical, consistent approach can lead to "a more valid and valuable" (p.31) type of research. This is because the researcher is, ideally, omni-present but organised, reflexive and continually endeavouring to address issues of accuracy and bias (Yin, 2011). Such self-reflection in the initial stages allows the researcher to interpret phenomena in natural settings and, in so doing, understand the social reality of the people being observed or, as in my case, interviewed. Field notes, compiled after each interview, helped me reflect the feelings of the participants. According to Snape and Spencer (2003), qualitative research should be flexible, sensitive to the researched and create a detailed re-presentation of the situation. As explained earlier, statistics can be objective 'facts', but they are static, cemented in time and place, whereas a re-presentation, which is as honest as possible, allows the reader to better understand social, dynamic phenomena such as employment, integration and racism (Butorac, 2014; Roshid & Chowdhury, 2013).

Another strength of a qualitative approach is that it can reflect multiple perspectives (Patton, 2001). In fact, in the data analysis stage "there are many threads that interweave through the complex set of interviews, reflections and observations" (Ezzy, 2002, p. 138). In taking such a bottom-up, inductive approach (Yin, 2011), it is hoped that the diversity of life can inform issues of interest to the wider public. For me, as for Roshid and Chowdhury (2013), this was achieved by interviewing participants from different age groups and backgrounds and who were employed in a variety of sectors, ranging from unskilled employment to tertiary education.

According to Patton (2001) and Yin (2011) a further advantage of a qualitative approach is that it aims to embody rigour and trustworthiness. By adding the author's voice and triangulating perspectives (Patton, 2001), the situation being experienced is recreated and made alive through the "linguistic turn" (Ezzy, 2002, p.149). Life is dynamic and social and, as poststructuralists claim, all that we know, our ontology, is shaped by the "partial, situational,

historical and provisional nature" (Ezzy, 2002, p.150) of experiences. Scientific research is partially shaped by the author although this is rarely acknowledged. The qualitative approach, on the other hand, is said to be more honest because it includes the author's involvement and aims to empower the researched while taking any political or moral issues that are being addressed into account. It was impossible to exclude myself from the research since I have been involved in language teaching for over thirty years and language testing for twenty-two years. Humans are not all the same and each person, each culture experiences phenomena in different ways and at different times. For this reason, qualitative research can be more time consuming and more difficult than quantitative methods but can ultimately reveal the "quintessence" (Ezzy, 2002, p.162) of the experience being lived by the 'Other'. The experience in this study is English language test-taking experience and employment of migrants from a variety of nationalities. Ultimately, fairness in depiction (Patton, 2001), compassion and understanding are the goals of qualitative research, rather than facts.

3.3 Limitations of an interpretive qualitative approach

One of the limitations of a qualitative research approach is that positivist theorists may claim it is subjective and therefore unscientific (Patton, 2001; Snape & Spencer, 2003). Indeed, quantitative research generally presents findings which are clear and can be easily utilised by lay people as well as various authorities. Reliability may also seem to be lacking when undertaking qualitative research since there is not one defined way to proceed (Patton, 2001; Snape & Spencer, 2003); rather there are multiple approaches which depend on the context and types of questions being researched. Patton (2001) points out that the resultant flexibility, while being an advantage, can also be misunderstood, especially if a variety of perspectives and interpretations are possible. The result can be a generic type of research which lacks a sense of quality (Snape & Spencer, 2003) with data which are not transparent and easily read (Yin, 2011). In addition, there are implications for the researcher when including her/himself in the report because the researcher is not anonymous. This means revealing their own struggles with possibly delicate issues. Ultimately, "Including personal experience is still not accepted by all social researchers "(Ezzy, 2002, p.153) so the researcher must decide if her/his own experience adds to the 'richness' of the research. As I was analysing and reflecting on the research data, I came to the conclusion that my own voice and experiences as an IELTS teacher and examiner were part of the interview process, so they could not be excluded from the study.

Further, Ezzy (2002) claims that the implications of the data in a qualitative study may never be thoroughly re-presented (Ezzy, 2002). This is because when writing, it is necessary to leave out certain details; the decision on what to include/omit can have serious implications for the research outcomes. Patton (2001) alleges that a qualitative approach necessitates "astute pattern recognition" (p.553) and implies that beginner researchers can lack experience with this and, hence, credibility. Issues that needed to be borne in mind when undertaking this research included how objectively/subjectively such research should be written and how politically charged it should be, especially when, for instance, issues of racism or the misuse of language tests through government policy become apparent in the research data.

3.4 Rationale for using an interpretive qualitative research approach for this study

Over ten years ago, Lumley and Brown (2005) stated that "In the last decade or so…..there has been a move from positivistic research focusing on properties of tests and scores toward a broader and more critical examination of a wide range of validity issues embracing construct definition as well as language testing practice and policy" (p. 833). This means that for nearly 25 years, language testing has evolved and started to embrace the social and societal consequences of testing systems in test validation arguments. Lumley and Brown (2005) claim that the issues related to high stakes language tests and language policies "necessarily rely on argument and the complexity of multiple points of view, and raise problems that challenge a positivistic paradigm and which no degree of quantitative sophistication can resolve" (p.834). For this reason, a qualitative research paradigm was deemed to be the most appropriate for my research into the consequences of language testing for test takers and test users in a New Zealand context.

With language testing, test score results may be much loved by bureaucrats since they epitomise the attraction of having a unidimensional 'truth'. However, as Shakespeare famously declared ''All the world's a stage". McNamara (2000) utilised the same theatrical analogy by comparing the test taking experience to an actor under a spotlight. The actor's abilities and psychological state are viewed in isolation at a particular moment in time (the 'old' approach to language assessment). Yet, the actor is part of a bigger picture which includes fellow actors, stagehands, producers, directors, the audience and critics, all of whom inform the wider public. Likewise, a test score result (the IELTS score in the case of this study) is co-created by a number of players, starting with the test creators who develop the tests, write the rubrics and set the standards, followed by the test takers and testers (examiners) who are engaged in a collaborative dialogue, trying to present and assess language

proficiency, which is a dynamic trait. The test users (employers and policy makers) create their own co-constructed meanings of the name of a language test such as IELTS or TOEFL (an acronym which they often cannot define) and the ensuing test score results. This is illustrated by Merrifield (2008; 2011) who reports that New Zealand organisations tend to take a score result because it has been endorsed by their Australian counterparts, thus co-creating a minimum standard without necessarily consulting with language testing researchers or any reflection regarding what can be inferenced about a test-taker who has gained (or not) a prescribed test score result. Numbers, such as an IELTS test score of 5.0, 6.0 or 7.0 can then take on a life of their own in the form of a disembodied number or 'truth', belying the notion that language proficiency is multidimensional and dynamic.

Snape and Spencer (2003) argue that the purpose of qualitative research is to counter this notion that there is only one 'truth' by taking an emic approach to the lived experiences of various phenomena in order to 'verstehen' [*understand*] and reflect the multiplicity of '*truths*' and '*values*' that permeate society through the eyes of the 'observed' or 'researched'. They stress that the crucial aspect is not only to judge but, rather, to reflect their perspectives by portraying a rich picture of participants' linguistic, but also affective or paralinguistic responses as these encapsulate the participants' feelings towards what they are experiencing. Consequently, the spontaneous affective responses of all test takers in this study are included in the findings, together with the interjections in the interviews made by one group, the New Zealand employers.

The movement away from a positivist paradigm, where truth is represented by statistics, has been reflected in the language testing arena with a movement away from test reliability towards test validity. Further, the concept of validity was initially defined in psychometric terms by looking at the constructs being measured and analysing how accurately a test measures what it is supposed to measure. In the last few decades, however, the focus of language testing researchers such as Lumley and Brown (2005), Kunnan (2018) and Spolsky (2008) has moved towards a qualitative analysis of the validation process. As Davies and Elder (2005) explain:

> The 'new' validity question is interested in what the test stands for, what it means, what inferences may be drawn from the scores derived from it. The 'old' version thinks it knows, the 'new' does not. For this reason, the 'new' offers an inductive approach. (p. 801).

Another change in language testing research has been the development of validation arguments. These are described by McNamara (2000) as being similar to a legal defence in court since they can be used to uphold or rebut claims about a test. The perspectives of multiple stakeholders are used in the validation process to shed light on the validity of a test through exposing stakeholders' views of the test's purpose, what they infer about test scores and the impact or consequences of the testing system in a particular context. This is explained by Davies and Elder (2005), thus: "Validity is not tucked up in the test itself but rather resides in the meanings that are ascribed to the test scores. These meaning are dynamic rather than static, changing along with shifting attitudes and practices in the relevant context of test use" (p. 809). Ultimately, language testing has become social and the societal impact needs collaboration between policy makers and language testing researchers to prevent unintended detrimental consequences.

Qualitative research aims to be useful for society with the intended audience being the wider public. As Roshid and Chowdhury (2013) conclude, they hope their research will be useful for "educational policy planners, teacher educators, teachers, employers and career advisers" (p. 79). Likewise, one of my aims was to provide data for policy makers and to help bridge the gap between test developers, language testing professionals and test users by highlighting the need for collaboration between all parties involved so that the complex nature of language assessment can be more fully understood. As Spolsky (2008) emphasised in his 25-year review of language testing, bureaucracies want "clear, unidimensional scores that can be quickly and easily interpreted for categorical decisions about the fate of test takers" (p.301). Yet, this ignores the fact that language proficiency is not static (Pilcher & Richards, 2017), that scores can change over time and that human abilities are multidimensional. In recent years, Chalhoub-Deville (2016), Inbar-Lourie (2008), Kunnan (2005; 2018), Lo Bianco (2018) and Spolsky (2008) saw the significance for language testers becoming involved in the debate about public policy and spreading language assessment literacy amongst relevant stakeholders.

3.5 Research Setting

As pointed out by Heidegger the research setting (cultural, historical, geographical and political) can have a distinct influence on the outcomes of the research. This study was initially undertaken in one of the main cities of New Zealand but as the research progressed non-native test takers and employers from across New Zealand contacted me to express their interest in the use of language testing for migration and they were, therefore, included in the study. As a result of the Canterbury

earthquakes in New Zealand (2010/2011) a large number of international workers were recruited in employment fields, ranging from carpentry and building to IT, project management and architecture. Meanwhile, as exemplified by the skilled migrant shortage list (INZ, 2018), the agricultural and manufacturing sectors across New Zealand were employing overseas sheep shearers and production workers due to a shortage of local employees. At the time of writing, the educational sector in New Zealand was reporting that there were not enough teachers for the start of classes in February 2019. For this reason, the New Zealand Ministry of Education was seeking to recruit teachers from overseas. Prior research into the use of IELTS had focused primarily on health practitioners (O'Neill et al., 2007; Read & Wette, 2009; Rumsey et al., 2016; Sedgewick et al., 2016; Wette, 2011) and, for this reason, I made the decision to investigate different employment fields where migrants were often employed. Thus, for this study, it was decided to focus on three sectors: education, IT/business and manufacturing/retail.

3.6 Data generation

The ways in which evidence for this study was gathered is outlined in Table 2. The reason for asking native speakers to take a sample IELTS test and subsequently to interview them about the experience was to provide data for RQ1: their perceptions of the relevance of the assessment and assessment tasks to the New Zealand domain. This was supplemented by the perceptions of New Zealand employers and non-native speaker test takers who had all been working in New Zealand for a number of years. The native speaker and non-native speaker test takers, together with the New Zealand employers were asked to give their views on the extent to which they believed the test tasks and constructs being tested reflected communicative language used in the New Zealand workplace (RQ3). The three groups of participants listed below were then asked how well they felt they could interpret and use the test score results (RQ2) and what they felt were the consequences of using language tests for employment in New Zealand (RQ4).

Table 2. Research phases, participants and data collection

Research phase	**Participants**	**How evidence gathered for research**
Phase One	Native speakers of English	Took a sample IELTS test which included the speaking, listening, reading and writing modules. Received grades for the listening and reading modules

		Looked at sample written responses written by non-native speakers in a published book Interviewed for their perspectives on the test taking experience Interviewed for their perspectives on language in the test compared with language in the workplace
Phase Two	Non-native speakers of English	Had taken an IELTS test in the past Worked in New Zealand Interviewed for their perspectives on the test taking experience Interviewed for their perspectives on language in the test compared with language in the New Zealand workplace
Phase Three	New Zealand employers	Employed migrants Looked at a sample IELTS test Looked at four sample written responses (two written by native speakers from Phase One and two written by non-native speakers in a published book Interviewed for their perspectives on the IELTS test Interviewed for their perspectives on language in the test compared with language in the New Zealand workplace

3.7 Testing and interviewing the participants

The three groups of participants for this research were native speakers of English (NSs) who identified that English was their first language, who worked in New Zealand in three occupational sectors (education, IT/business and manufacturing/retail) and who volunteered to take a sample IELTS test. The second group were non-native speakers of English (NNSs) who had taken the IELTS exam and now work in similar occupational areas. The non-native speaker group came from a variety

of countries where they learned English as a second language. They had all needed to take the IELTS exam for immigration and work purposes when coming to New Zealand. They had been employed in New Zealand for lengths of between three and twenty years. The employers recruit overseas people in the same employment fields and were interested in language assessment for employment in New Zealand.

Participants were initially recruited via flyers and emails to schools, universities, employers and businesses who recruited and/or employed a variety of nationalities. The recruitment process snowballed as interested students, employers and organisations forwarded the flyers on to their members, employees and colleagues. After the first set of interviews with an employer, and two native speakers who took the test, it was decided to select participants more purposively in a similar approach to other research which has investigated IELTS results and job placement (Al-Issa et al., 2016; Roshid & Chowdhury, 2013). This was because the first three participants all mentioned that, in their opinion, an IELTS type of test would be impossible for young people as they would not have the required linguistic skills. Thus, it was deemed important to include a range of ages, genders and occupations in Phase One (NSs), a range of nationalities, ages, genders and occupations for Phase Two (NNSs) and a corresponding mix of employers (Phase Three) who employed native and non-native speakers in the three main employment sectors: education, IT/business and manufacturing/retail. I sought to include an equal mix of genders in case males and females had different views or test score results. This did not lead to any meaningful outcomes, although that could not be anticipated at the beginning of my study. A person who was a health and safety specialist was purposively included in Phase Three due to the fact that the majority of the people interviewed mentioned that employees needed adequate reading and writing skills primarily to understand health and safety regulations at work.

One of the first findings was that many native speakers were unwilling to take a test of their own English language proficiency, so much so that Phase One was the most difficult area to recruit for. A number of potential participants contacted the researcher and professed interest in being tested/interviewed, only to withdraw a week or so later. For this reason, data gathering took approximately one year, with Phase One taking six months, starting in September 2017 and finishing in February 2018. The difficulty with data collection in this phase was exacerbated by the Christmas holiday period, a time when it is often difficult to contact people in NZ as the summer school holidays last from the end of December until the beginning of February. The native speaker participant test takers who did take part either worked with migrants or were genuinely interested

in language testing. The details of the participants are outlined in tables 3, 4 and 5. They were categorised into males and females in case gender proved to be important in the data analysis.

Table 3. Phase One participants: Native speakers of English who took a sample IELTS test

Participant	Age bracket	Occupational area	Occupation
F1	60s	Education	Primary school teacher
F2	50s	IT	ICT manager
F3	40s	Retail	Retail assistant
F4	20s	Hospitality	Waitress
F5	40s	IT	Accounts manager
F6	30s	Education	University student – Master's level
M1	40s	Education	University lecturer
M2	60s	IT	Telecom technician
M3	20s	Retail	Supermarket worker
M4	60s	Education	IELTS/ESOL teacher
M5	20s	Education	University student Undergraduate who had part-time jobs in IT and retail.
M6	Late teens	Education	High school student – Year 12 who had started an apprenticeship.

Phase Two was conducted between February and May 2018. As Gribble et al. (2016) discovered, it is often useful to use a combination of purposive and snowball sampling to gather empirical data, and that was the case with this study. Non-native speakers were eager to pass on the research details to friends and colleagues so participant recruitment for this phase snowballed relatively easily. The researcher was vigilant in maintaining an even balance of genders and occupations in case these proved to bear any significance to the findings.

Table 4. Phase Two participants: Non-native speakers of English who had taken an IELTS test

Participant	Country of origin	Age bracket	Occupational area	Occupation and length of residency
F GNNS	Germany	30s	Education	Primary school teacher. 15+ years in NZ.

F MNNS	Malaysia	40s	Education	ESOL teacher and PhD student. 3 years in NZ.
F PNNS	Poland	30s	Customer service	Librarian. 20 years in NZ.
F FNNS	Philippines	40s	Medical science	Laboratory scientist. 5+ years in NZ.
F VNNS	Vietnam	40s	Retail	Retail assistant. 3 years in NZ.
M GNNS	Germany	40s	Education	University lecturer. 15+ years in NZ.
M SNNS	Singapore	20s	Education	University student – Undergraduate. 3 years in NZ.
M CNNS	China	30s	Retail/logistics	Businessman. 15+ years in NZ.
M INNS	India	30s	Manufacturing	Production manager. 5 years in NZ.
M FNNS	Philippines	40s	Manufacturing	Manufacturing worker. 5+ years in NZ.

The participants in Phase Three of this study, the employers, were interviewed between May and September 2018. Bodies such as the Chamber of Commerce and the Retail Association of New Zealand sent out emails outlining the research to their members, which helped with recruitment since I was only contacted by employers who employed a variety of nationalities. Most of these employers were aware of the fact that migrants to New Zealand had to undertake some kind of language testing as part of the immigration process, but the degree of their knowledge varied widely. They were generally interested in the research and viewed the interview sessions as a type of professional development, a factor recommended by Barbour (2007) as being beneficial for professional people who are participants in research.

Table 5. Phase Three participants: Native speaker test user employers (all in their fifties and sixties)

Participant	Gender	Occupational area	Occupation
Emp 1	M	Education	University employer
Emp 2	F	Education	Senior educator
Emp 3	M	Education	Primary school principal
Emp 4	F	Education	ESOL director
Emp 5	F	IT	IT recruitment director
Emp 6	F	Business	Business consultant
Emp 7	M	Business	Immigration business director
Emp 8	M	Manufacturing	Head of recruitment
Emp 9	M	Retail	Retail business director
Emp 10	F	Manufacturing	Health and safety consultant

3.7.1 Testing native speakers of English

In order to ascertain if the test scores are used in valid and ethical ways in New Zealand, it was decided early in this research that testing native speakers of English would be appropriate, since determining results from native speaker New Zealanders who are employed in the three sectors under investigation (education, IT/business and manufacturing/retail) should identify if decisions made about the test score results of non-native speakers are being made ethically. The test score results from any language test cannot be deemed to be used ethically or appropriately if non-native speakers of a language are required to gain higher scores than the average native speaker, and as Kunnan (2018) outlined, the unethical or unfair use of language tests has occurred in many countries. In discussion with my supervisors at the outset, it was decided that authorities would heed the language testing outcomes of native speakers more than non-native speakers (who are expected to struggle). I had initially hoped to interview native speakers of English who had taken an IELTS test. IELTS allowed me to put flyers in the test centre without the IELTS logo. However, no native speaker candidates came forward. For this reason, I sent flyers to a variety of institutions and employers throughout the Canterbury region asking for native speakers who would volunteer to take a sample IELTS test. The reason behind asking native speakers of English to take a sample test was to explore if they believed this type of language test is an appropriate, valid and meaningful tool to check the language ability of employees in New Zealand as stated in RQ1: How well do test takers and test users perceive the relevance of the assessment and the assessment tasks to the New Zealand domain? Their perceptions were also relevant to RQ3: How well do test takers and test

users feel that the IELTS tasks and the constructs being tested reflect communicative language in the New Zealand workplace? Since native speakers of English take the IELTS test for a variety of reasons, it is important to analyse their perceptions of the consequences for native speaker test takers (RQ4).

3.7.2 Choosing a sample IELTS test

I contacted IELTS for a sample test and they advised me to use the latest IELTS test book from the Cambridge University Press series, which was *IELTS 11* (2016). The front cover states that the book contains authentic examination papers. The tasks were chosen at random with Task One (Appendix L) being a description of pie charts, Task Two (Appendix M) a discursive essay about learning languages, and the two-minute monologue for the speaking test (Appendix N) being a description of a writer the test taker would like to meet. The academic version of the test was chosen because this is what is required by bodies such as the New Zealand Education Council for teacher registration although Immigration New Zealand will accept both the Academic and General Training Modules. Some readers of this research may be concerned about the use of the Academic version of the test since this is not needed for residency. However, the fact that the immigration authorities accept both tests, as well as First Certificate in English (FCE), Occupational English Test (OET), Pearson Test of English (PTE) and Test of English as a Foreign Language (TOEFL) indicates that there may well be a lack of language assessment literacy amongst test-users regarding the similarities and differences between these two tests. It is commonly thought by language teachers that I have taught with that a pass in FCE (First Certificate in English) denotes lower English proficiency than IELTS 6.5 and that the Academic version of IELTS is more difficult than the General Training test. Nonetheless, there is little evidence for this and indeed many stakeholders are not aware that the speaking and listening sections of the test are exactly the same, while for the writing section Task One is different but Task Two is similar (a discursive type essay). The Task One writing differs in that candidates have to describe a graph/graphs for the Academic test and write a letter in an appropriate style for the General Training test. However, the scripts are marked against the same rubrics (Appendices O and P).

Stakeholders are often unaware that in the reading section, candidates need to score 23 out of 40 to gain a 6.0 band score result for the Academic test whereas they need to gain 30 out of 40 to receive the same score, a 6.0, on the General Training test (Appendix Q). The assessment rubrics are exactly the same for the speaking and writing sections for the Academic and General Training modules while the listening and reading sections have the same question types and can be marked by people who are not trained to be language testing raters. It can, therefore, be argued that the Academic and General Training versions of IELTS have similar difficulty levels.

3.7.3 Phase One test taking process (NSs)

The research participants in Phase One took a sample IELTS academic test, which included the speaking, listening, reading and writing sections in that order, to be as similar as possible to an actual testing day. A limitation of this research is that the native speaker test takers did not take an actual IELTS test. However, the participants took the test-taking process seriously, strict time limits were adhered to and they were reluctant to put their pens down at the end of each section. The speaking section of the test was not graded, mainly because IELTS examiners are not allowed to allocate test scores to prospective test candidates for ethical reasons. Likewise, Phase One participants did not receive grades for their writing; rather they were interviewed about the test taking experience. It was presumed that the native speakers of English would gain high scores in the speaking module, and since the objective from the outset had been to undertake qualitative research, the perceptions of participants towards the speaking and writing modules were deemed to be more important than their scores. However, Phase One participants did receive scores for the reading and listening modules. This was possible because *IELTS 11* contains acceptable answers for the reading and listening sections of the test. The writing scripts were given at a later date to two experienced examiners, who had been examiners for 15+ years. They assessed the written scripts independently, and these scores were used in the form of a table (table 6 in section 4.4.1) as prompts for participants to comment on in phases two and three of this research.

3.7.4 Phase Two data generation (NNSs)

Phase Two participants (NNSs) had all taken an IELTS test in the past and now worked in New Zealand. Some had taken the test once while others had taken it multiple times. While some non-native speaker test takers presented me with their test score results, and others could remember their test score results exactly, some discussed only the range of scores they had received because they had taken the test more than four times. The NNS participants were asked to reflect on their experiences with language testing and compare this with the type of language they now use for working or studying in New Zealand. A comparison of their results with those of native speakers, and their perceptions of the language skills being tested, were used to help elucidate whether test users are using test score results in valid and ethical ways.

3.7.5 Phase Three data generation (Employers)

Phase Three participants (employers) were questioned regarding their knowledge of language assessment. They were shown sample tests and were asked to comment on how the language in the test compared with language used in their workplaces. The test scores of the native speaker

volunteers (Phase One) were organised into a table to compare with the results that the non-native speakers of English (Phase Two) reported having received. Since I had samples of test responses for the writing part of IELTS, these were used for the employers to reflect on the type of language that was produced in their fields of employment. Two writing responses from Phase One were used as prompts. These were a Task One response (graph description) written by the university lecturer (M1), and a Task Two response (discursive essay) written by the accounts manager (F5). These native speaker writing responses were complemented by two sample non-native speaker writing responses published in *IELTS 11* (2016), together with examiners' comments on the writing scripts of the non-native speakers. The employers were then asked to compare the language used in the sample responses the language they would expect to see in their workplace, and if their perceptions of language testing had changed. This was particularly relevant because test-users, such as employers, are not experts in language assessment and yet they may use the test results as part of a recruitment process.

3.7.6 In-depth interviews

Since most participants were working, it was necessary to conduct qualitative interviews with each person, either in their workplace or at a place that they preferred. The setting was, therefore, convenient for the participants because it was chosen by them and this allowed them to speak freely in a one-on-one situation. The interviews with participants took around 45 minutes to gain in-depth perspectives. The participants were able to express their thoughts, practices and beliefs in a non-threatening way while I, as the researcher, could react to the dialogue and use prompts to uncover issues that appeared to be significant for the person being interviewed. A stereo IC recorder was used to ensure quality recordings. The recorder was highly sensitive and had automatic level adjustment for voice recording, together with a direct USB connection for ease of transfer of files.

The interview questions were approved by the university Human Ethics Committee (Appendices A and B) and trialled with one native speaker employer and one native speaker test taker. Asking employers to view a sample video of a speaking test was omitted because this took too long. These questions, shown in appendices I, J and K, followed the following general format:

- What did you think about language testing and what would be included?
- After viewing a sample test have your perceptions of language testing changed?

This line of questioning relates to the relevance of IELTS to the New Zealand domain (domain inference) as posed in RQ1: How well do test takers and test users perceive the relevance of the assessment and the assessment tasks to the New Zealand domain?

- Looking at these examples (writing tasks), which ones do you feel are competent users of English? What scores do you believe a competent L1 speakers would receive?

These questions relate to how well test takers and test users can interpret and use test score results (decision inference) and follows on from RQ2.

- How does the test content and constructs being tested compare with the English that is used in your workplace?

This question relates to whether test takers and test users feel that the IELTS tasks and the constructs being tested reflect communicative language used in the New Zealand workplace (RQ3).

- Do you have any comments for authorities who use language test results for business, migration, and professional registration purposes?

This final line of questioning led the research participants to comment on RQ4: What do test takers and test users perceive to be the consequences of test use for the New Zealand context.

The interview questions were semi-scripted to allow for conversations between the researcher and participants which varied depending on the participants' backgrounds (Stevens, 2017). Since knowledge is co-created, interviews are a beneficial way to collect data because they allow the researcher to gain detailed insights of the views, experiences and beliefs of participants in an emic approach (Snape & Spencer, 2003). The purpose of utilising semi-structured interviews was to allow for the collection of rich data, and as Gill, Stewart, Treasure, and Chadwick (2008) discovered, it was not necessary to compile too many prompts in advance as questions can sometimes "take on a life of their own" (p.294). For this reason, between seven and fourteen pre-determined questions were used in each phase of the study (Appendices I, J and K), and these were changed depending on the responses of the interviewees. For example, if a participant seemed to be concerned about cultural or employment issues, these were explored in more depth. Due to their one-on-one nature, it was also possible in the interviews to investigate sensitive issues that participants may not have wished to discuss in a group environment. As recommended by Gill et al. (2008), the interviewer started with questions that could be answered easily to build rapport before moving on to more complicated issues. Each question was followed by probing remarks and clarification techniques to ensure the views of participants were fully explored. A hermeneutic approach suited both the collection of data from semi-structured interviews and the research

question since issues related to fairness, equity and validity need to be handled sensitively and probed using follow up questions which respond to the participants' assertions. Janesick (2003) compares qualitative research with the arts of choreography and dancing, explaining that research should initially follow fixed movements similar to those in a minuet but as the research progresses, the researcher is more likely to improvise, following on from discussions with participants.

The interviews generated rich data and followed Foddy's Symbolic Interactionist Model where messages are encoded by a participant and decoded by the researcher (Wengraf, 2001). Symbolic Interactionism focuses on how individuals actively construct meaning by interpreting and taking into account the actions of others (Barbour, 2007). A broad approach to Symbolic Interactionism can also relate interaction to social, economic, political and policy contexts, which is particularly useful as the interpretations and uses (in policies) of IELTS band scores may have political overtones. In my study the social, economic and political consequences of language testing were all mentioned.

For hermeneutic conversation analysis, it is important for the researcher to listen, respond in a respectful, tactful way and guide the conversation while allowing topics that are important to the participants to be addressed. During the interview sessions, it often appeared to me as if participants were off topic, but I gave them time to air their views, some of which proved to be invaluable later. As Gill et al. (2008), point out one advantage of semi-structured interviews, compared with structured or unstructured interviews, is that they allow the researcher/s to discover values, attitudes and/or beliefs that are salient for the participants but were not anticipated by the researcher.

I initially intended to conduct focus group sessions with Phase Three participants (employers) to encourage a lively debate amongst decision makers. This proved impossible as busy employers are generally time poor and dates that were convenient for one employer were not convenient for others. However, the individual interviews with the employers provided rich data reflecting different cultures and attitudes towards language in the three areas of employment. This led to the collection of a myriad of perspectives which could be compared and contrasted in a hermeneutic circle.

3.8 Use of Software

The interviews were initially read and re-read many times to gain an intimate relationship with the voice of the speaker and to try and enter his/her world. Initial codes were identified at this point, as recommended by Castleberry and Nolen (2018) but in order to gain an analysis that was as robust as possible, NVivo 12 software was used. NVivo is particularly useful for qualitative data analysis because it can help the researcher code and retrieve units of text, and thereby improves the researcher's ability to think conceptually and place these units within an appropriate theoretical framework (Braun & Clarke, 2006). However, Castleberry and Nolen (2018) note that "While software can assist researchers with organizing large amounts of qualitative data, the researcher's mind is the power behind analysis and not any software programme" (p. 809). More importantly, using computer software can lead the researcher to be distant from the data (Braun & Clarke, 2006) so the original scripts were constantly referred to in order to retain the essence of the participants' voice. This constant moving from the original texts to the data stored in the software package reflected the hermeneutic approach of moving in circles; repeatedly analysing the parts to eventually build a picture of the whole (Langridge, 2007).

The main advantage of using NVivo 12 was that codes could be generated and changed as the participants' viewpoints challenged each other for relevance and prominence. Memos and explanations, which were attached to the codes, aided the organisation and retrieval of data. The coding system was further developed when new codes, such as the educational status of test takers, which had not been previously considered were created and when some codes became redundant or overlapped with each other (Braun & Clarke, 2006).

3.9 Hermeneutical data analysis

As discussed in 3.4, qualitative research in general seeks to re-present phenomena experienced, felt and articulated by people who have participated in, have knowledge or an impression of an event or situation. The researcher should endeavour to appreciate, analyse and reflect the participants' perceptions in an inductive way, which means moving looking for patterns in the data to develop broader concepts of their views of the test content and test taking process. Contextualisation is an important factor in interpretive hermeneutic inquiry. The broad context for this research is the linguistic repertoire that is generally needed in the contemporary New Zealand workplace in three employment groups (education, IT/business and manufacturing/retail)

3.9.1 The position of the researcher

I have taught and assessed students' English language proficiency for more than thirty years, and my position as an IELTS examiner for 22 years is a factor which may influence the outcomes of this research. As noted by Ezzy (2002), Snape and Spencer (2003) and Yin (2011), the presence of the interviewer cannot be erased from the research because the traditions of both the interviewer and participants affect what and how they communicate. For this reason, hermeneutics seeks to understand prejudice (not in a negative sense but with an understanding that we all pre-judge or pre-conceive situations), the spoken word and tradition (semantics, etymology and culture are apposite terms here) as well as understanding as a way of being. For the above reasons, the researcher's background and possible prejudices (Stevens, 2017) as a teacher and examiner were fully explained to the participants prior to commencing the interviews, and to the readers of this research in the introduction to this study. While an examiner's or teacher's predilection for noting breadth and depth of morphosyntactical and lexical complexity in language, together with issues related to cohesion and coherence, were sometimes referred to by participants as well as the researcher in the interviews, the participants' reflections on what constitutes language competence in a normal workplace took precedence. Similarly, my own knowledge that language testing does not necessarily mean that native speakers gain scores as proficient users of their own language, as highlighted by Stricker (2004), was always in the background but at the same time this was not the focus of the data analysis; rather the reaction of the participants towards high-stakes language testing compared with their own or their colleagues' English was paramount.

Secondly, when interviewing it was necessary to take into account how research is socially constructed in the interactions between researcher and participants and to remember that the cyclical hermeneutic process consists of "repeated and emerging reflection on the data" (Mackey, 2011, p. 106), interpreting the viewpoints of the participants (Snape and Spencer, 2003) in a dynamic path of discovery (Taylor & Bogdan, 1998).

3.9.2 The viewpoints of the research participants

The participants in this research were IELTS test takers and people who employed migrants (test users). Kunnan (2018) outlines potential stakeholders for language assessment studies, ranging from school teachers to university professionals but, as he states test-takers "are the centrepiece of the assessment enterprise and hopefully the beneficiaries of assessments. Finally, there is the group that is arguably most powerful of the whole lot: institutional administrators, government officials and policy makers" (p. 225). Language testing can be considered ethically appropriate and the test itself

valid if stakeholders believe they can understand and interpret test score results accurately and reliably. This means the test score results need to make sense to all stakeholders, including test-takers and test-users. The hermeneutic approach was used in this study to seek understanding of the experience of taking a high stakes language test (the phenomenon) by test-takers and how test results are viewed by the wider New Zealand community.

Taking a hermeneutic approach aligns with my epistemological view of the nature of 'knowing' and 'being'. Such an approach entails using participants' perspectives to 'Verstehen' [*understand*] (Snape & Spencer, 2003) phenomena, to re-present their 'Dasein' [*their feelings or sense of being when they were present*] (Heidegger, 1967), and to interpret the data collected using a cyclical or spiral approach with the aim of understanding the whole through an analysis of the various parts and how these parts relate to each other. Stevens (2017) claims that hermeneutic interpretation is not only a practical approach to research analysis but is also "an ontological structure of understanding" (p. 72). By using hermeneutic or interpretive phenomenology as a researcher, I have sought to thoroughly understand and interpret "the essential features, hidden meaning, and essence of a conscious experience" (Al-Issa et al., 2016, p.852). Understanding the view of the 'Other' is important because each research participant has their own reality and ways of understanding that reality. In my research the 'Others' are the research participants who represent language testing stakeholders in a New Zealand context. The hermeneutic process that was used for my data analysis is outlined in Figure 2.

3.9.3 The hermeneutic circle or spiral

According to Gadamer (2014), the researcher needs to have some connection to the phenomenon being researched and to undertake a circular or spiral analysis of the data that have been collected as outlined in figure 2. This means reading, taking note of initial interpretations, and revisiting the data on numerous occasions to uncover new meanings that are emanating from the text. This iterative process was fluid and ongoing and I was consciously aware that initial interpretations can be misleading and that a deeper understanding of the viewpoints of people, from different sectors of society and different cultures with their own traditions and histories, can only emerge slowly by linking the parts (of their conversations) into a coherent whole. For this reason, the interviews from each phase of the research were read and then laid aside while the researcher ruminated on initial interpretations and their meanings for the relationship between language testing and linguistic competence in society. When using the hermeneutic circle, analysis should move back and forth between the data and research to form new understandings of the perspectives of the research participants. The aim is to find confirmations, contradictions and new considerations (Mackey,

2011). At the same time, the researcher needs to be 'suspicious' of any biases or possible distortions that may eventuate (Mackey, 2011; Stevens, 2017).

The research context is affected by language, culture, rituals, political structure and history (Cerbone, 2012). In their research into the use of IELTS as a language test for Arab English language teachers, Al-Issa et al. (2016) perceive the IELTS test as being a "social, educational, pedagogic, linguistic and cultural phenomenon" (p.853) which can be interpreted "constructively, reflexively, reflectively, creatively, and innovatively to generate rich data" (Al-Issa et al., 2016, p. 853). The multiple perspectives of a range of people working and studying in New Zealand reflect the participants' reality at a certain point in time and in a specific place. An analogy can be made here with language assessment which is a snapshot of certain language skills on a particular date. The language testing results do not reflect stable traits as they can be influenced by, amongst other variables, topics covered in the exam as well as affective factors on the day. Yet, test-users may misinterpret test results, often presented in numerical form, as being a factual, finite representation of a candidate's language competence.

Figure 2: The hermeneutic circle or spiral

1. Preunderstanding

2. Hermeneutic circle of interpretation

3. Fusion of horizons. Researcher able to re-present new understanding of phenomenon with the aid of participants' perspectives

3.10 Thematic analysis

Thematic analysis can be used to compile and analyse data in logical steps, and for this reason, this approach was chosen to help develop the codes and themes and, subsequently, to interpret them (Castleberry & Nolen, 2018) using a hermeneutic approach. Reflecting on the similarities and differences within and across groups and allowing new themes to emerge could be termed a Grounded Theory approach. However, Braun and Clarke (2006) advise against using the term Grounded Theory as an overarching term because it is often over-used, leading to grounded-theory 'lite' analyses, which are not directed to theory development. Braun and Clarke (2006) contend that the term "named and claimed" (p.8) thematic analysis is preferable, namely because it can mean a more approachable form of analysis, one which does not need detailed and technical knowledge of what constitutes Grounded Theory. Castleberry and Nolen (2018) claim that, to be thorough and rigorous, thematic analysis comprises five steps: compiling, disassembling, reassembling,

interpreting and concluding (Castleberry & Nolen, 2018). These are the steps that I employed in this study.

3.10.1 Data compilation

Barbour (2007) argues that interviews can be time consuming but that they allow the researcher to include salient quotes from individuals. Using direct quotes gives a voice to the participants (Bogdan & Biklen, 2007) so vignettes which highlighted the essence of the experience of taking a language test and noteworthy reactions to this type of test were collected in memos in NVivo. These quotes are included to help the readers understand the contextual narrative and to show that the story truly reflects the "quintessence" (Ezzy, 2002, p.162) of the participants' experiences.

The recorded interviews were audio-recorded and transcribed verbatim by me, which means that the initial coded sections were not de-contextualised and the presence (Dasein) and feelings of the participants were taken into account. As Braun and Clarke (2006) claim, transcribing interviews manually can be "time-consuming, frustrating, and at times boring" (p. 17). However, for Castleberry and Nolen (2018), transcribing the interviews are a way to bring the researcher closer to the data since it allows the researcher to re-familiarise themselves with the data and can be seen as being part of an interpretive approach to data analysis (Braun & Clarke, 2006). The main advantage for me was that I was continuously reminded of the context and the participants' reactions when transcribing because their voices were being heard repeatedly.

Field notes were made directly after the interviews and during transcription to record the feelings that aspects of language testing evoked in the participants' faces and through non-verbal utterances such as laughter (indicating surprise or other feelings) and semantic choice (Wengraf, 2001). In these notes I tried to capture the authentic mood of the interviews through noting paralinguistic features such as intonation, points of emphasis or surprise and expressions used to convey feelings such as 'wow, ugh, great'. For analysis purposes, Wengraf (2001), explains the importance of using the lexicon of the participants as well as taking paralinguistics into account since meaning is not only reported in words.

3.10.2 Interview analysis – disassembling the interviews

For the analysis of the transcripts I began with the cognitive process of frequently reading, re-reading and reflecting on the data as advised by Braun and Clarke (2006) and Castleberry and Nolen (2018). This continual reflection allowed the researcher to become intimately connected to the

underlying ideas, categories and themes being expressed in the transcripts and to disassemble them. Braun and Clarke (2006) explain that thematic analysis begins with identifying codes (idea units) or elements of the raw data that appear to the researcher to be related to the research question. Codes can be grouped into categories when relations between the codes are identified. Themes or units of analysis, on the other hand, are broader. They are meaningful groupings of categories which have some connections with each other. Identifying these themes allows the researcher to begin interpreting the data.

Initial data analysis produced the following themes: 1) the participants' feelings about this type of language assessment, 2) the issues (or not) they perceived they had with the IELTS test, 3) how the test is related to competency in the workforce/society, 4) how the average New Zealander would perform on this type of assessment and 5) the advice they have for decision makers (test-users) who use the test results. The native speakers' remarks were further categorised into their perceptions regarding the four skills (reading, writing, speaking and listening) and how essential each skill is for workplace competency.

After immersing myself in the data (Braun & Clarke, 2006; Mackey, 2011) by re-reading the interview scripts, the five categories above were changed, and ten themes emerged. According to Braun and Clarke (2006), it is advisable initially to code for as many themes as possible because it is not usually clear at this stage which themes are most relevant. The ten themes at this stage were: general feelings about language testing, perceptions of test content, clarification questions about how language is assessed, the scores and their meaning, relevance to NZ work/study/society, views on their own performance, views on non-native speakers of English and their test results, the relevance of educational/professional status, test performance issues and impact on New Zealand society. This was enlightening because after recording the interviews, I was unaware that Phase One participants had mentioned the educational/professional status of test-takers. My only recollection was that this had been discussed in Phase Two.

Phase Two participants (NNSs) were interviewed regarding their IELTS journey, their feelings about this journey and their beliefs about how this compares with language competence in their current workplace. It was expected that there would be different stories from a variety of cultural backgrounds although this did not prove to be the case. The initial themes that emerged were similar to the Phase One participants (NSs) but with an added emphasis on their language testing journey and perceptions of their own language competence.

Finally, data from Phase Three participants (New Zealand employers) were analysed. From the first reading of their transcripts, themes of importance to this group appeared to be their lack of knowledge about language testing, and the non-relevance of language testing results for the workplace. Coffey and Atkinson (1996) recommend identifying and re-ordering data, accompanied by reflection so that the different perspectives of participants can be discovered. When taking a heuristic, contrastive approach, the exceptional cases can also prove to be important (Coffey & Atkinson, 1996). For me, what proved to be unexpected was the concern from employers about low literacy levels in society.

Data from the interviews were analysed inductively (Castleberry & Nolen, 2018; Patton 1990) by identifying codes, categories and emerging patterns. Castleberry and Nolen (2018) explain how different approaches to coding can be established: "In Vivo" coding (p. 809) uses the actual words and/or phrases used by the participants, descriptive coding uses basic topics referred to in the data, open coding means the coding emerges as analysis is developed while a priori coding can use coding utilised in research on similar topics. Castleberry and Nolen (2018) recommend using a combination of these coding methods. By using a cyclical process, my initial coding could be re-analysed and refined. As Davidson and Tolich (1999) explain, the data will, in all likelihood, contain "complex and interwoven variables" (p. 22). For me as a researcher this meant being patient and flexible enough to react to changes in the narratives being told (Davidson & Tolich, 1999).

3.10.3 Reassembling

Themes were then created by placing codes into context with each other, that is into categories, and related to theory. Using thematic analysis, hierarchies were developed in order to tell a coherent story, as recommended by Braun and Clarke (2006). From a constructionist epistemological stance, I sought to interpret the semantic content of the data to identify underlying ideas, assumptions and conceptualisations, taking into account the socio-cultural contexts in which the discourse was produced (Braun & Clarke, 2006). Coffey and Atkinson (1996) recommend identifying and re-ordering data, accompanied by reflection so that the different perspectives of test-users can be discovered. Al-Issa et al. (2016) described how they used a cycle, moving from texts to codes and codes to theory in order to capture the underlying meaning of the interviews which "is anchored in the deep structures of the interview texts" (p.855).

For me, the transcripts had to be continually re-read and re-examined to establish connections within and across groups, and to refine these connections in light of the theoretical framework, as recommended by Castleberry and Nolen (2018) and Mackey (2011). Following on from constant

reflection of the Interpretation Use Argument (IUA) approach as analysed and described by Kane (1992, 2002, 2006, 2013, 2016), Bachman and Palmer (2010) and Kunnan (2018), the codes and ensuing categories that started to emerge from the data were broadly related to the following inferences: test domain, extrapolation, decisions and consequences. Although the IUA approach was in the back of my mind when relating the codes, categories and themes to each other, I began by organising the participants' narratives into vignettes which represented their reflections on the IELTS test at an individual level, followed by their perceptions of the impact of the test at a societal level.

Reassembling the data entailed repeatedly analysing the interviews. The first themes that appeared to be important for all interviewees were the participants' knowledge of high stakes language tests, their perceptions and affective responses to the test content, and their reactions towards the skills being tested in relation to the language used in their workplace. The participants also commented on the test scores, if they understood them and how test scores related to educational level. At this point, I tentatively ascribed these themes as being evidence for the domain and decision inferences (how the test is used in a New Zealand context and whether diagnostic results are useful for employees and employers to make decisions about language competence), and extrapolation inference (whether the test is testing language skills/tasks commonly needed in the New Zealand workplace). In addition, it seemed that the data reflected the participants' general awareness of IELTS and the traits (educational level) they believed were being tested.

The second set of themes were reassembled into vignettes which portrayed the participants' reflections on the societal impact of language testing for migration purposes. A wide range of issues were mentioned which I realised would be useful as evidence for a consequence inference (whether the test has consequences which are beneficial to test-takers and test-users). In this sense, a combination of clean slate and grounded theory approaches were utilised (Stacey, 2002) in a thematic analysis approach. According to Castleberry and Nolen (2018) coding can stop when no new themes become apparent when the researcher is reviewing their data (saturation point). This occurred when the range of societal impacts were reassembled.

3.10.4 Interpreting

Interpretation of the data was ongoing throughout the analytical process but as Yin (2011) points out, it is important to make this clear by stating how the interpretation started, continued and ended for the final narrative to appear complete. At the outset, I accepted that the messages conveyed by the participants were 'correct' (Stevens, 2017). I then needed to interpret these messages in a way that came naturally but with "tact, understanding and openness" (Stevens, 2017,

p. 83). Being open means the interpreter can develop and change their own perspective of the phenomena being discussed and in so doing reach a fusion of horizons (Gadamer, 2002). Hermeneutic inquiry can be 'messy' (Stevens, 2017), but the interpretation is persuasive (Mishler, 1990) when it is insightful (portraying multiple perspectives), and practical when linked to the research question or issue in society which engendered the research in the first place (Stevens, 2017). As data reassembling and interpretation continued in this study, it became apparent that two issues were being raised by my research participants which corresponded with my research question; how valid native speakers perceived an international language test as being, and whether they believed using a language test was treating test takers ethically. The interpretation of research data should be fair and credible (other researchers would interpret the data in a similar fashion), accurate through representing the raw data well, and add value to the debate. As I was interpreting the perceptions of the participants, the social and economic consequences of high stakes language tests such as IELTS were becoming clearer to me. Rigour and trustworthiness are examined in more depth in 3.13.

According to Castleberry and Nolen (2018), themes are salient if they capture something important and can lead to global findings across all the codes. The issue which led to my interest in the social implications of language assessment were the situations faced by overseas teachers seeking registration with the Education Council and professionals/business people seeking to reside permanently in New Zealand. Interviews with employers highlighted the fact that immigration policies are just as much a problem for them trying to recruit and retain staff as they are for the test takers.

3.10.5 Concluding

The final stage in Castleberry and Nolen's (2018) five-step thematic analysis process is concluding. Once the data in this study had been interpreted, the findings were closely related to the research question to try and bring all disparate parts together into a whole for the conclusion. As pointed out by Castleberry & Nolen (2018) and Stevens (2017), it is often the case that the research question has changed over the course of the research. My research questions were originally:

- Research question one: How do the tasks of the IELTS test relate to the language needs of Filipino and other migrants who are seeking employment as skilled migrants in Christchurch, New Zealand?
- Research question two: How is IELTS used in a New Zealand context for migration and employment?

- Research question three: How are the IELTS band score requirements which have been stipulated by different authorities in reading, writing, speaking and listening perceived by these test takers and what is the impact on the community of needing to work towards these scores?

These were adapted to the following two questions:

- Research question one: Is performance on the IELTS test a reliable predictor of success in the N.Z. workplace and what are the views of test-takers as regards their language competence on the job vis-à-vis the band scores they have attained.
- Research question two: To what extent are IELTS test scores used in valid and ethical ways for the purpose of immigration, employment and professional registration?

These were then narrowed down to the second question:

- To what extent are language test scores used in valid and ethical ways for the purposes of immigration, employment and professional registration in a New Zealand context?

The following four sub-questions were used to analyse the perceptions of the research participants:

RQ1: How well do test takers and test users perceive the relevance of the assessment and the assessment tasks to the New Zealand domain?

RQ2: How well do test takers and test users feel they can interpret and use test score results?

RQ3: How well do test takers and test users feel that the IELTS tasks and the constructs being tested reflect communicative language used in the New Zealand workplace?

RQ4: What do test takers and test users perceive to be the consequences of test use for the New Zealand context?

When concluding, Braun and Clarke (2006) state that it is easy to take a naïve approach to thematic analysis by "giving voice" (p.7) to research participants. While the extracts of data that are chosen to be presented should be "vivid" (Braun & Clarke, 2006, p. 23), the ensuing story is not coherent and compelling unless it is accompanied by an explanation of what is interesting and why that is so. Consequently, the final report needs to provide the reader with "a concise, coherent, logical, non-repetitive, and interesting account of the story the data tell – within and across themes" (Braun & Clarke, 2006, p. 23). Consequently, in the narrative I endeavoured to not merely describe the data but also to make an argument in relation to the research question (Braun and Clarke, 2006).

3.11 Using the research questions for test validation

According to Braun and Clarke (2006), researchers must continuously bear in mind their theoretical frameworks when compiling and interpreting their data. As they point out "What is important is that the theoretical framework and methods match what the researcher wants to know, and that they acknowledge these decisions, and recognise them as decisions" (Braun & Clarke, 2006, p. 8). Accordingly, the essential aspect of taking a hermeneutic approach for my research was to present an interpretation of how language assessment is experienced by native speaker and non-native speaker test takers by using an argument-based approach towards test validation. As the data compilation and interpretation evolved, I continued to look for evidence of the ethical implications of using IELTS for migration and whether my research participants believed test score results reflect workplace competence.

Drawing on Chapelle et al's (2010), Chapelle, Cotos and Lee's (2015), Johnson and Riazi's (2017), Kane's (2013) and Kunnan's (2018) conceptualisations of inferences and principles, supported by claims, warrants and backing or rebuttals, I wanted to show what IELTS means for test takers and test users in New Zealand, whether the test has consequences which are beneficial for New Zealand society and whether the research participants believe decisions being made on test performance by professional bodies and immigration are appropriate for the New Zealand context. In this sense, I have used the research questions to explore and interpret the research participants' views of test validation. Critical hermeneutical analysis was used to reveal issues related to fairness and justice (Kunnan, 2018) in language assessment and language assessment policy. The key objective at all times was to ascertain if test-takers and test-users perceive that language assessment test scores are used in valid and ethical ways for the purpose of immigration, employment and professional registration in a New Zealand context.

3.12 Data analysis summary

The four research questions were used to explore issues related to the validity of using high-stakes language tests for living and working in New Zealand. Using a hermeneutic approach involved a continual cycling between the data, the background theory and the four sub research questions. Similar research questions may be created to address the individual, group and societal impacts of language tests in different contexts.

3.13 Ethical considerations

A key ethical principle should be to "do no harm" (Tolich & Davidson, 1999) to the researched and, yet, all research can potentially do some amount of harm. Therefore, it is crucial only to research issues that are "theoretically valid and socially significant" (Tolich & Davidson, 1999, p.71). Issues related to employment and migration are of considerable importance and can have life-changing effects, as demonstrated by Roshid and Chowdhury (2013) and Gribble et al. (2016). Nonetheless, it is crucial to remember that the researchers are in a position of power and typically gain the most from research (Bogdan & Biklen, 2007). My involvement with IELTS as an examiner already implies a position of power, which needed to be addressed from the outset in communication with participants. For this reason, the participants were given advance information about the study, the background of the researcher and assurance about ethical issues, such as anonymity and confidentiality

The participants were contacted by phone, Skype or email to ensure they were willing and able to be interviewed. This was especially crucial with Phase One participants because the testing, followed by an interview, took approximately four hours per person. Participants were sent the interview questions and information sheet that had been approved by the university's Human Ethics Committee in advance (Appendices C- H). The research intent and ethical considerations were reiterated prior to the beginning of each interview to ensure participants fully comprehended the scope of the research.

Anonymity was not possible since the participants' identity became apparent to me during the research, but confidentiality could be promised by using pseudonyms/no names or other distinguishing information. This was reinforced at the start of each interview session when participants were given oral and written guidelines about the contact details of people to approach if they had any reservations about the research. Any volunteers need to have informed, educated consent which includes the option of withdrawing at any time and to understand that they should not feel coerced by others in their community to participate (Cohen, Manion & Morrison, 2009). One educator withdrew from the research because they were worried that the publication of any research would hurt their professional registration status and career prospects while one of the test-users asked for their job title to be more generic, in order to facilitate greater anonymity.

In the recruiting and interview stages, it is also advisable to have flexible principles that suit the context and evolve with the research process. This means considering the 'Relationships

Paradigm' as well as the 'Principlist Paradigm' (Cullen, 2005). These principles include respecting people, being fair and right (justice) and maximising benefits (beneficence) while doing as little harm as is practically possible. At the same time, it is crucial to acknowledge the changing context and relationships between people, including the relationships between researcher and researched. During the interview sessions, participants appeared relaxed and did not shy from voicing their opinions while the researcher took care to make sure that participants who took the IELTS test for research purposes (Phase One) did not feel demeaned in any way. The fact that most test-takers in this research (Phase One and Phase Two) felt that other New Zealanders would have similar test results (or indeed lower results) to them indicates that they were not overly worried about what the results showed about their literacy skills. It must be acknowledged that all participants were employed in New Zealand and that the results did not threaten their work, professional or immigration status.

Additionally, it is important to always treat participants and their cultural values with respect (Bogdan & Biklen, 2007) even if, as Hara (1995) claims, that cultural differences mean it is not possible to be totally value-neutral. Since Phase Two participants came from a range of countries, including India and South East Asia, it was necessary to take into account that research participants may respond in a way that aims to please the interviewer since the interview is, in effect, a social practice (Talmy, 2011) co-constructed by interviewer and interviewees. Talmy (2011) reported that interviewees adapted their responses according to the identities they wished to convey and, in this way, they re-produced their language ideologies in and through the interview itself. This was a factor to take into account in my research, not only considering the participants' country of origin but also the socio-economic status of participants as revealed by their area of employment. It was not possible to ascertain if any participants did give responses that they felt would 'look good' but they did acknowledge that their language was not 'perfect' and were keen to point out that this did not hinder them in the workplace.

One key problem is that New Zealand has a small population base and can be described as a village where people often know each other and, therefore, the identity of test takers and test users may be easily recognised. This appeared to be more of an issue for test users (Phase Three) so they were assured that the names of the companies they worked for would not be used, only general terms such as manufacturing employer or education employer. Having said this, Merrifield (2011) stated that some of the professional decision makers in her study would have liked to have had the opportunity to try the IELTS test and, in fact, most native speaker

test takers and employers in this research looked on the interview as a type of educational exercise. The use of language testing for migration purposes was of great interest to some of them because they needed to employ migrants and some had spent time training employees, only for them to then be denied work or residency visas. During Phase Three the employers were asked to comment on the results of native speaker candidates and on their written output, which could have involved some ethical risks. However, this was avoided since the native speaker candidates were not named and could not be identified.

After each interview, the participants were asked if they had any other comments to make and were promised a summary of the research findings. This was to ensure that participants felt they were contributing positively to the research and that their voice had not been misrepresented. Ensuring that culturally safe practices are maintained is not only important in the data gathering phase but also with regards to publication, especially as IELTS is a high stakes exam. Thanks to technology, information is ubiquitous; "data misuse, therefore, or disclosure at the wrong time or to the wrong client or organ" (Cohen et al., 2009, pp. 71/72) could have serious outcomes. Therefore, participants were fully aware of the time involved, the amount of time data would be kept, and that research may be published.

3.14 Rigour and trustworthiness

For credibility purposes, it is important to remember that the researchers' values cannot be detached from qualitative interviews since every person brings their identity, values and beliefs to an interview situation. However, Harrison, MacGibbon and Morton (2001) point out that the researchers' beliefs can prove to be advantageous since "Through judicious use of self-disclosure, interviews become conversations, and richer data are possible" (p.323). According to Taylor, Bogdan and DeVault (2016), credibility and trustworthiness are strengthened by including the researcher's theoretical framework, frame of mind and how that changed, together with the author's relationship with the participants. For this reason, I was keen to establish a professional relationship with the participants and to emphasise my role as a language teacher and examiner. At the same time, my field notes and memos in NVivo contained the participants' spontaneous responses and feelings as portrayed by intonation and/or facial gestures together with my reactions to their statements. According to Davidson and Tolich (1999), adding the author's voice and triangulating perspectives, the interpretation of the data becomes more honest and valid because the findings that are re-presented "accurately reflect the opinions or actions of the people in the study" (p. 28). This allows the reader to truly understand the nature of reality of a particular community and how

they ‘feel’ complicated issues, such as employment, integration and possibly racism (Roshid & Chowdhury, 2013; Butorac, 2014; Gribble et al., 2016).

Another issue discussed by Bogdan and Biklen (2007) is to try and discuss reservations that the reader may feel when reading any interpretations of the data. As much as possible, I have tried to pre-empt the readers’ possible concerns so that the account of the perceptions of the need for language testing for migration to New Zealand is sensitive to the multiple realities and experiences of those being researched. The relevance of using the Academic as opposed to the General Training IELTS test has already been discussed in 3.7.2. Another aspect of this research that needs to be highlighted is that the Phase One participants (NSs) volunteered to sit a sample IELTS test, and therefore the results for them were not life changing in any way. For this reason, it could be argued that they took the test in a casual way and did not prepare adequately. However, each participant in Phase One took the test very seriously and they were reluctant to finish until they felt they had answered each task as competently as they could. In addition, some Phase One participants researched the test on-line while one participant, the IELTS teacher, could be described as having superior knowledge of the test constructs.

Coffey and Atkinson (1996) confirm that empirical evidence needs to be gathered but they remind researchers to be wary of drawing simplistic conclusions in order to ensure rigour is achieved. They also urge researchers to constantly ask themselves if the questions being addressed measure the phenomena they are researching. In the operationalisation stage, qualitative research is made stronger through triangulation (Davidson & Tolich, 1999; Cohen et al., 2009), or crystallisation (Richardson, 1994) by looking at phenomena (people and events) from different angles and by using different methods. For this reason, I was constantly self-reflexive, taking the perspectives of different test-takers and test-users into account and comparing the ensuing data. The research results were strengthened by including the test results of native speakers of English, their perceptions of the test, together with the data collected from test-user interviews. The fact that the perceptions of native speaker test-takers or test-users towards these high-stakes language tests have attracted little research to date means that native speakers’ perceptions of the linguistic repertoire actually used in the modern New Zealand workplace should add to the debate of the current validity of language tests what were originally designed in 1988. Throughout, I have attempted to present the findings in a clear, comprehensible way so that they should ultimately help test users better

understand the meaning of IELTS band scores when they are used for migration to English speaking countries.

3.15 Summary

This chapter has described the qualitative research paradigm and how this suited the study of the interpretation of language testing for migration to New Zealand. In addition, the research setting, the research participants and the reasons for taking a hermeneutic approach to data analysis have been outlined. A systematic approach to data analysis was undertaken to lend trustworthiness and credibility to this study.

The perceptions of the research participants on a personal or individual level, followed by the group or workplace level and finally the societal impact, will be presented in the next three chapters. This echoes the call by Chalhoub-Deville (2016) to explore intended and unintended interpretations, uses and decisions of language assessment results in the individual, aggregate, and larger educational-social contexts. The impact of the test at the individual level will include the participants' awareness of testing systems, their reflections on the appropriateness of language tests for migration, the relevance of test-content for their workplaces, their perceptions of test difficulty and their interpretation of test score results (data for RQ1 and RQ2). The impact of high stakes language testing at the group and societal levels will include the research participants' views on how language is used in today's workplace and how this compares with the test content and constructs being assessed (data for RQ3). Finally, the consequences for New Zealand society will be discussed (data for RQ4).

CHAPTER 4

FINDINGS PART ONE: PERCEPTIONS OF ENGLISH LANGUAGE ASSESSMENT AT AN INDIVIDUAL LEVEL

4.0 Introduction

In this chapter, key themes from interviews with native speaker research participants (test- takers and test-users) are discussed on a personal or micro level. This includes their views of what this type of language test signifies in terms of test content and what the test scores meant for them individually. The analysis included the participants' knowledge of this type of test prior to seeing a sample IELTS test and their initial responses after finishing or viewing the test. This is compared with the perceptions of non-native speaker test-takers who are working in New Zealand. The themes which emerged during interview analysis were: prior knowledge, initial perceptions, how participants viewed what is being assessed (the test content), how the test is assessed (what can be inferred about test scores), and their own performance. This included performance issues such as timing, test-taking strategies and affective contributors to test performance. These perceptions provided evidence for RQ1: How well do test takers and test users perceive the relevance of the assessment and the assessment tasks to the New Zealand domain? The ability, or inability, of participants (native-speaker test takers, non-native speaker test takers and New Zealand employers) to interpret test scores appropriately helped build a validation argument for RQ2: How well do test takers and test users feel they can interpret and use test score results? A further, non-anticipated, theme emerged because the participants voiced their opinions about the role of discrete item language testing focusing on accuracy and spontaneity (higher language cognition) compared with English for communication (basic language cognition).

RQ1: How well do test takers and test users perceive the relevance of the assessment and the assessment tasks to the New Zealand domain?

4.1 Knowledge of the test and test content.

In this section, the amount of awareness from participants (native and non-native speaker test takers and native speaker employers) about large scale language testing provided empirical evidence about the relevance of this type of test for New Zealand society. Secondly, the data helped elucidate if test content is appropriate for New Zealand stakeholders and can therefore be ethically and validly used for migration purposes.

All of the participants were aware that language testing for migration purposes existed and the majority had heard the term IELTS, although many were unsure what this acronym meant. However, the amount of knowledge that participants had about the test ranged from very high to negligible.

4.1.1 High level of knowledge

The participant with arguably the most knowledge of IELTS was the native speaker English language teacher (M4) who taught IELTS classes as part of his job. He was the only native speaker participant who had previously seen a high-stakes language test. This participant had previously worked for the New Zealand government as a parliamentary speech writer and felt confident in his own English language skills. For M4 (ESOL teacher), teaching IELTS consisted of three areas: helping students with their English language skills, helping them understand what he labelled as the mechanics of the test, followed by encouraging them to overcome affective factors which impede test performance.

The university lecturer (M1) was conversant with IELTS requirements for university entrance and, since he had international students in his classes, was aware of the band score results they had achieved and how these related to their academic performance at university. He was very confident in his academic language abilities, so much so that he argued that his responses in the reading test were better than those in the marking guide. One of the employers (business consultant, Emp 6) had worked with IELTS students at her local university, helping them prepare for the listening section of the exam and, for this reason, she was familiar with the test, especially the oral sections.

All of the non-native speaker participants in my research had taken the IELTS test at some time in the past and this group were, therefore, cognisant of the test content and, moreover, voiced their opinions on test-taking strategies. Their memories of taking the test differed since some had taken it recently as opposed to many years previously while some had taken the test only once compared with some who had taken it multiple times.

4.1.2 Average level of knowledge

The native speaker Master's level student (F6) worked with people from a range of nationalities who had taken IELTS. She was studying Tagalog and had a reasonable awareness of tests from her secondary and tertiary studies. Similarly, the native speaker primary school teacher (F1) used literacy testing systems with her students and knew from working in South East Asia not only that IELTS was a common testing system used worldwide, but also that people she met overseas talked about struggling to achieve a level 5.0.

The native speaker ICT manager (F2), accounts manager (F5), retail assistant (F3) and telecom technician (M2) had all heard of IELTS although they had no knowledge of learning a language and their own experiences with tests were based on high school exams which they had completed many years previously, before working in their chosen fields. However, they had heard of IELTS in a wider sense; their remarks included "it's an English test so foreigners can get jobs here or schooling here" (telecom technician, M2), and "Just that it's an English test that all foreign people have to do to be able to stay here" (ICT manager, F2). The telecom technician (M2) reported that he had looked for and studied IELTS strategies presented on-line prior to taking the test, mainly because he was a technical as opposed to a literacy person and was, for this reason, concerned about his language skills.

As for the employers, all of those in the education sector were aware that there were IELTS requirements for professional registration or for immigration purposes. Similarly, one of the business employers (Emp 7) was a director in the immigration industry and for this reason was knowledgeable about band score results; all of the non-native speaker employees in his company had needed to score 7.0 or above in IELTS for professional registration as immigration advisors if they were not New Zealand citizens or permanent residents. Likewise, the IT employer (Emp 5) and ESOL employer (Emp 4) were regularly presented with IELTS score results when people were applying for jobs, even though the companies did not request test scores as part of the employment process.

4.1.3 Negligible amount of knowledge

The younger native speaker participants (F4, M3, M5, M6) had not heard of IELTS, although F4 (waitress) did remark that "I just knew that you had to do an English test to get into a country or to be able to work in certain parts of a country or something like that". This group were least confident with their own English language with the youngest stating "It was hard for me because I don't read at all and I'm not the best at writing" (M6, 17-year old high school student), despite the fact that he was in year 12 at high school and taking NCEA (National Certificate of Educational Achievement) exams.

The employers in the manufacturing and retail sectors had very little knowledge of high stakes language testing for immigration purposes and yet they wanted information about the testing process because it was having an impact on their ability to recruit staff. Even though the manufacturing employer (Emp 8) had no knowledge of language assessment, he explained that "I

haven't really had anything to do with it, but I've certainly heard about the IELTS and we get a lot of migrants who specify an IELTS score or something that they've received" (Emp 8).

One of the business employers (business consultant, Emp 6) worked as a consultant with a wide range of employers in the Canterbury region. She explained the situation for employers in general:

> I would guarantee that almost every employer that I have spoken to has no idea what IELTS means. They understand that it's some sort of test, but they wouldn't understand the different levels and what it means for them. They just want their employees to be able to speak and understand good English and obviously for certain roles they need to be able to write and read it.

4.1.4 General perceptions

The majority of the native speakers recognised that the test was similar to English tests that they had taken at school and, for this reason, it had what can be termed as face validity for them. They recognised some of the question types and felt that it was important to test the four skills of reading, writing, speaking and listening, although they differed in their opinions of how important each skill was for integration in New Zealand. As for reliability, the university lecturer (M1) knew the name IELTS and felt it was the best exam of its kind "a necessary evil", not ideal but not subjective either so, in this sense it could be relied upon. Likewise, the ESOL teacher liked the rigorous nature of the test, asserting that the way that English is taught to EAL (English as an additional language) students for this type of test would be useful for all young people in the New Zealand education system.

The non-native speaker participants who were interviewed for my research were working in the areas of education, IT/business and retail/manufacturing in New Zealand. Some had held high level jobs for a number of years. The non-native speakers of English who had practised test-taking strategies through taking IELTS specific lessons felt the test met their expectations and, in general, felt it was a high-quality standardised assessment of their English abilities. In general, test stakeholders, such as employers and teachers of English for exam purposes, know that there are two versions of IELTS, and they tend to believe that the Academic module is more difficult than the General Training test. It is, therefore, informative to note the perceptions of test takers who had experienced taking both modules. Some of the non-native speakers who I interviewed had taken the General Training version of IELTS. They found no difference between the General Training and Academic versions in terms of the test content and the scores they had received with the two tests

being identified as equally difficult “I don’t think there’s that much difference between the Academic and General test. It’s basically the same stuff” (University student, M SNNS).

All of the employers interviewed for this study were native speakers of English who had little to no experience of learning a second language. Their awareness of the breadth and depth of knowledge needed for second language acquisition was, therefore, limited. Half of the employers, especially those in manufacturing and retail, believed that a language test would be assessing general communication. However, even those involved in education thought communicative competence was being assessed:

> I know there are assessments in reading, writing, speaking and in those areas to get a feel for what it is for someone who comes to New Zealand and their ability to speak, communicate, write and I suppose to be an active citizen in the new place they are coming to and not to be in a place of disadvantage (Senior educator, Emp 2).

One of the frequently mentioned advantages of IELTS was that it has an international reputation. For the senior educator (Emp 2), this meant that teachers who registered in New Zealand could be accepted by the education authorities in other countries. The business consultant had a similar impression to the university lecturer (M1) as she stated that “IELTS is the best of a bad bunch, I guess, because it’s international” (Emp 6). The immigration business employer (Emp 7) also recognised the benefits of having an international test since “tension around somebody’s English has now been contracted out so I don’t have to make a judgement on your English language skills, neither does Immigration New Zealand”.

Although the business consultant (Emp 6) could see the merits of IELTS because it shows employers that migrants have a reasonable grasp of English, she conceded that it is a “blunt tool” and results do not translate into communicative competence in a workplace setting. The widespread use of the term IELTS was generally seen as being positive, but the university employer (Emp 1) queried why he always heard the name of this language test as opposed to any others, indicating that the term had little value for him:

> Why this particular one, why has this become the benchmark that everything is measured against would be a question that I would ask because it’s always IELTS. We hear about it all the time...I don’t know why the Education Council have chosen that. It may well be almost a colonial throwback. That it’s come from Cambridge and it’s IELTS so therefore it must be fit for purpose (Emp 1).

4.1.5 Perceptions of test difficulty

As IELTS purports to test the language used in communication in English speaking countries, it may be assumed that native speakers who use English for communication in work or study arenas would be able to demonstrate competence in their mother tongue. However, all of the native speakers who volunteered to take a sample IELTS test, found that the test was much harder than they had expected. Comments included "I think it's quite hard. Well, it's hard for me as a native speaker" (Retail assistant, F3) and:

> I thought it would have been easier to do it. I just think it's unfair now because, like, it's just too hard because if people like me can't do it, like people who live here and speak the language. Why should people from another country be able to do it (Waitress, F4).

The employers' initial responses when presented with a sample test was similar to the native speaker test takers; all of them declared that it was much harder than they had anticipated:

> You know this is actually harder than I had expected it to be. You know there's a lot of vocabulary in this that… You could grab the average man off the street, and they would struggle with it, eh? Interesting. So, you need to go back and find the passage? I think that would be quite hard for anybody really (IT employer, Emp 5).

On top of this, the interjections some of the employers utilised demonstrates how surprised they were: the university employer's first reaction was:

> Wow, so what reading age is this pitched at?...Wow., the fact that it is time-bound as well so if you don't finish it you are going to score low… Oh, it's much, much harder than I anticipated, especially the writing part (University employer, Emp 1).

The senior educator echoed these views, repeating "Gosh, it's quite challenging. There's a lot of cross checking. I'd say that given the literacy work I've done in schools and even with quite senior students, lots of this would actually be quite tough." (Senior educator, Emp 2). The manufacturing and retail employers showed surprise at how the language in the test was unrelated to workplace language and unrelated to New Zealand life:

> Gosh, the language is pretty out there, isn't it? It's reasonably complex, like biodiversity for example. Not useful for my workplace, not at all…Gosh, that's quite full on (Manufacturing, Emp 8).

> And you're looking at how well they can understand their English, aren't you? So it was a Brit who wrote this because now you've got barges. I don't think this is for New Zealand. Woah, that's pretty heavy going isn't it? Not at all a common theme at all. Real broad general knowledge, not. I wouldn't have a clue about this – it just doesn't make sense at all (Retail director, Emp 9).

In contrast, all non-native test takers seemed to be content with the oral sections of the exam since they recognised how necessary it was to have communicative competence. The two participants from Germany who had not done so much test preparation, a primary school teacher (F GNNS) and a university lecturer (M GNNS), were surprised by how difficult the test was, and did not feel that the test results reflected their English language skills, especially in the writing section. In fact, the issue for nine out of ten of the non-native speaker participants was that their scores for the writing section were lower than those for the other skills. This resulted in a range of emotions, including being scared about taking the test, feeling depressed, or losing confidence in their English abilities. As the German primary school teacher (F GNNS) explained:

> I studied literacy for three years and got A+'s in literacy assignments ...I felt quite down at one point thinking I had picked the wrong profession to be in but then I work at three schools and they were all absolutely appalled about me having to do an English test.

Likewise, the university student from Singapore (M SNNS) had received A grades for his assignments at university but scored lowest on the writing section of IELTS and, for this reason, had had to take the test multiple times. This confused him since he regarded himself as a native speaker of English. The librarian who was originally from Poland (F PNNS) had taught English in Japan. She found the test taking experience "painful, very expensive and full of frustration". Strong emotions were also felt by the Filipino and Malaysian communities. The ESOL teacher from Malaysia (F MNNS) commented that taking IELTS is "very scary" and that most of her compatriots were fearful of taking IELTS because the test results can have life changing consequences in terms of migration. The librarian (F PNNS) was so affected by her writing score that for many years she had avoided places or times where she needed to write. Yet, she had taken a New Zealand Diploma in Human Resources and passed a written entry exam to become a librarian. The Filipino scientist was dismayed with her writing score because she had won journalism prizes in the Philippines for articles written in English while the Malaysian PhD student and ESOL teacher (F MNNS) had received training from an IELTS examiner at the school where she worked and was, at the time of interview, able to write well enough to undertake a PhD in English. Altogether, this led to a lack of confidence in the

meaningfulness of the writing section and to doubts about whether the test score results could be extrapolated to academic and vocational domains.

Interestingly, the female participants from the Philippines (F FNNS) and Malaysia (F MNNS) had a negative experience from receiving 7.0 for their writing while the Indian production manager (M INNS) and the Vietnamese retail assistant (F VNNS) gained in confidence from receiving a lower score, 6.0 and 6.5 respectively in the writing section. These two participants from India and Vietnam reported having received some of the lowest overall band score results out of the non-native speaker participants but they both felt a good IELTS score was very important for social integration and working in New Zealand.

4.1.6 Summary

Nearly all the participants had heard of IELTS indicating that it is relevant to the New Zealand context. However, apart from the non-native speaker test takers, there was very little awareness of what is included in the test. The international reputation of IELTS added to its meaningfulness. Initial responses demonstrate that when the native speaker participants viewed a sample test, it transpired that this type of test was much harder than all of them had anticipated.

4.2 Assessment tasks: Test content

The participants' views on the skills being tested was useful to help shed light on one of the key claims made by the test developers that "IELTS ensures the right level of English to succeed in study, training or life in a new country" (ielts, 2019). The perceptions of native speakers of English, non-native speakers of English and New Zealand employers regarding the test content in relation to the TLU (target language use) domain is divided into two sections: oral skills (speaking and listening) and written skills (reading and writing).

4.2.1 Assessment tasks: Receptive and productive oral skills

The aspects of the test that native speakers regarded as being most relevant to New Zealand life were the speaking and listening sections. Participants valued the communicative nature of the speaking section "Well, I think being able to speak and communicate orally is useful so that is fine... the questions were a bit odd, but whatever" (University lecturer, M1). The first and third section of the speaking module were commented on most favourably because they were conversational in nature, although some declared that the topics themselves and the frequent change in topic made these sections seem unlike a normal conversation.

The second section, where candidates are asked to speak on a topic for two minutes, led to problems for most of the native speakers. In particular, the Master's level student (F6) and the high school student (M6) found this section 'daunting'. They were asked to describe an author they were interested in, a typical prompt for this part of the IELTS test. The issues with the two-minute monologue were created by the topic and the length of time they needed to speak. Many reported that due to work or study pressures, they never read books and could not speak for a whole two minutes on a topic they had no interest in: "Yes, but when you ask me to tell you about a book I read, I think Christ, what was the last book I read?" (ICT manager, F2), "I ran out of things to say. I think I had covered everything I could think of. I mean I don't know that much about many authors, I don't know a lot of detail" (Accounts manager, F5). As the university student (M5) perceptively explained:

> I know that in theatre sports there's a game where it's similar to that. You have to act on something but it's usually act on the colour blue for one minute. It's not like talk about something that's very specific so that two minutes seemed like a long time because no-one would, well very few people would know a lot about a writer, like enough to talk about them for two minutes... You don't very often talk for that long, for two minutes solidly, by yourself.

The telecom technician (M2) pointed out that the construct of spoken fluency was intertwined with knowledge of a specific topic "I found the fluency one quite tricky mainly because it's mixing the content of the author with the ability to talk non-stop for two minutes". Age may have been another factor which played a part in the monologue as the youngest participant (M6, high school student) could not respond at all, indicating that this participant would need help with test taking strategies, even though he felt he was a competent speaker of his mother tongue.

As for the listening module, most of the native speaker test takers found this relatively easy with the Master's level student (F6) commenting that "The listening section was the one I probably scored the highest in and that was easy because I could understand exactly what was going on and it was clear and it was clear what was being asked" while the ESOL teacher (M4) stated that this section was "surprisingly clear". Other participants, however, were sceptical about the constructs being tested with the university lecturer (M1) wondering how picking out specific facts as opposed to inferring meaning when listening to a whole conversation was an adequate test of his listening skills. This reflects research by Aryadoust (2013) who argued that the IELTS listening measures only two constructs: understanding explicitly stated information and paraphrases rather than the ability to infer meaning from long texts or to understand functional, everyday language.

Commenting on listening skills in a workplace environment the accounts manager (F5) pointed out that "If I put it in a real-world situation, I would listen to something, like, in a meeting and I would take notes whereas to fill in words, when would you ever do that? How real is that?" while the university student (M5) had a similar viewpoint "I can't think of any situation other than a test where you would have to listen to a presentation and wouldn't be allowed to go over what they've shown or said". Half of the participants found the section where they needed to label a map difficult, stating that they had never listened to a presentation and labelled a map at the same time. The listening context was another area which attracted criticism with a few participants commenting that it should be more New Zealand specific "One of them was more associated to the English countryside which I didn't think was relevant to New Zealand, just seemed quite weird" (Retail assistant, F3).

All of the non-native had similar opinions to the native speaker test takers since they regarded the listening and speaking sections of the test as being of utmost importance for living and working in New Zealand. Unlike the native speakers though, they did not generally find the questions in either the speaking or listening modules unusual. This is presumably because these types of questions, which appear to be inauthentic to a native speaker of English are prevalent in practice activities in foreign language classrooms.

The ESOL teacher from Malaysia (F MNNS) regarded the speaking and listening sections as being 'authentic' and extremely useful for living in New Zealand:

> And I think they've done a good job, the IELTS, in the sense of so many skills and things like that. I find it very useful in terms of listening because when I came here, that's how people are, you know. It's very authentic. I guess, yes, I think it's a good measure to see whether someone is able to speak good English or not.

Commenting on the relevance of the listening and speaking sections for the New Zealand workplace, the Filipino manufacturing worker (M FNNS) pointed out that work skills, followed by communication skills are paramount for success in the New Zealand workplace: "I believe the most important is our skill, second is the communication. And that's why we are here - to work". (M FNNS). While the primary school teacher from Germany (F GNNS) identified speaking, followed by listening, as being necessary skills to be successful as a teacher, she argued that the listening test "was totally out of context because it was British and there was British people speaking to you and they say things a bit different although they didn't have a strong accent".

The Vietnamese retail assistant (F VNNS) emphasised the need for oral communicative competence because she needed to be able to communicate well with her customers and gaining a 6.0 for the speaking section of the test was beneficial for her in this regard. Even so, she was still scared speaking to local inhabitants outside of work, indicating that a score higher than 6.0 in the speaking and listening modules may be advantageous for social inclusion.

The New Zealand employers' perceptions were no different to those of the test-takers; they prioritised oral skills over written skills for their living and working in New Zealand. A typical response was:

> In terms of working collaboratively and actually being able to fit into the society and being able to assimilate and make bonds and form connections, it's the speaking and listening that makes it (IT Emp 5).

Although it may be assumed that written skills would be a priority in education, oral skills were most crucial for employers in this sector. The responses of the primary school principal (Emp 3) echoed the opinion of the German primary school teacher (F GNNS) as he emphasised that relationship building (with children, their parents and other staff members) was the most important skill for teachers. The senior educator (Emp 2) reiterated this, saying:

> Obviously critical is listening and speaking equally – around those relational aspects with students because it is that connecting piece and that sort of daily contact, connecting… as teachers the bulk of your work is listening, speaking, understanding, sort of interpreting ...The listening and speaking are so immediate and for the teacher in the classroom, something that you don't pick up readily can snowball and escalate into things and can move very quickly and if it's repeated over time, it has a real impact on the learning and the ability of the teacher to teach (Emp 2).

Likewise, as the university employer (Emp 1) pointed out, good spoken English is paramount in order to command respect, give instructions and to communicate with students.

For the manufacturing and retail employers, the ability for staff to liaise with customers as well as other staff members was crucial. Cultural differences were mentioned more than language skills because employees needed to be confident to voice opinions, check for understanding and ask for clarification if unsure. Using workplace jargon, as well as being polite and comprehensible were considered necessary for productive oral competence. For receptive competence, listening to short, sharp instructions was noted as being paramount as opposed to listening to lectures. Although

listening skills were highly rated, employers did mention that generic listening skills were not enough; rather, understanding the New Zealand accent, New Zealand slang and how indirect New Zealanders are in the workplace was essential. Surprisingly, cultural differences also had an impact on employees from different English-speaking countries according to the IT employer (Emp 5). This was because, in her experience, Americans could not understand New Zealanders' indirectness while New Zealanders found Americans rude and this led to tensions within this global IT company.

4.2.2 Oral skills summary

There is strong backing from these research participants that a good performance in the oral sections of IELTS may have a connection with employability. The sections of the speaking and listening modules which assess dialogic competence were most valued. Nonetheless, a stronger New Zealand element would have been preferred.

4.2.3 Assessment tasks: Receptive and productive written skills

The main points raised by the native speaker participants (both test takers and test users) about the reading section were related to the topics, the lexical density of the passages and the constructs being tested through the question types.

First of all, looking at the test constructs, the university student (M5) felt that the reading passages and questions were similar to what was necessary for tertiary studies "reading articles and getting little bits of annoying information out. I use that all the time because I'm at uni". However, the university lecturer's (M1) tone was indignant about his reading score, suggesting that his answers were better than those in the answer schedule:

> No, that's not right. Let me have a look at that. Those are right and that's definitely right. Let me have a look at that. These are because there are specific things you are looking for? …But I put central axle and then crossed it out when I realised they wanted only one word and then I put arms. Oh, that's ridiculous… (University lecturer, M1).

All of the test takers involved in education commented that the reading section focused more on following instructions and finding specific lexical items rather than overall comprehension. The primary school teacher (F1), used the same adjective to describe the reading section as the university lecturer (M1) as she found it "ridiculous". In her experience with school tests it was important to look at the clear intent of the answer:

> If you are trying to test their understanding of English you should be looking at the intent, the clear intent of the answer. You are not marking their English ability here, you are there

> for assessing their ability to follow instructions… I really felt the questions, many of them had been intentionally written to trip people up (Primary school teacher, F1).

Additionally, the university lecturer (M1) commented on the specificity of the prescribed answers "The reading part was again very specific, detail focussed so you don't get the gist of the communication, you don't get the overall picture...But generally speaking, that is not important for anything I could imagine, that level of exactness" while the telecom technician (M2) reiterated these points, saying that:

> You know like looking for the author's name and then looking for the reason so there's more structural and if you're logical and well organised - it's not necessarily a test of your language ability. It's also highly reliant on your reading speed. There were some questions that I reckon there were several answers to… I'm sure there was – the one with the shaft and the wheel and the axle.

Likewise, the retail assistant (F3) felt that she had understood the essence of the reading texts, but her answers may have been incorrect due to the prescriptive nature of the questions:

> And I think that for me, I would have spelling errors and like I put two words, but the question asked for one word so what I put was actually right but because I put two words it was wrong. So, I'd actually got the communication but because of the instructions ...two words would actually make more sense for me than one because it wouldn't be descriptive enough.

Relating the specificity of the answers to the workplace, the Master's level student (F6) who worked as a manager in a manufacturing company argued that "We're not in the workplace and pretty much in business and life in general, you're not picking one word out here – that's not picking out meaning from anything" (F6). She suggested that a test of overall comprehension would be better than a test which asks for specific lexis.

As regards the topics and lexical density of the reading passages, the test takers who were office managers commented that the kind of topics in the reading section prevented them from being engaged in the task:

> I guess I do read a lot of information and I pick out the bits that are relevant so yes, but I think it was the topic I struggled a little bit with … I thought oh God this is boring. The first one was kind of interesting, but it was hard though. I think it was harder than a workplace or

> the stuff I read in a normal workplace…not a lot of people need to read screeds in the workplace (Accounts manager, F5).
>
> The reading for me – I didn't like the topic. I didn't like having to find and answer the questions because I didn't have an interest in the topics and there was just SO much information in the articles (ICT manager, F2).

As explained in 3.7.2, the participants took the Academic version of the IELTS test and in the particular test that was chosen, the texts tended to be scientific and technically orientated. However, migrants to New Zealand can take the Academic module for permanent residence and need to do so for professional registration.

The non-native speaker test takers were more appreciative of the reading module than the native speakers of English although they did comment on the topics being out of their range of interest. For example, the librarian from Poland (F PNNS) explained that:

> Also, there are some tasks which are completely out of interest. For example, there's a reading instruction task which is about the built of the engine of a boat. Come on, what's my interest about the engine of the boat? If it was, for example, about cooking or making a nice cake or sewing, which I'm very into, oh yeah, I would just go for it and I would achieve a high score. But for people who don't know sewing at all and have no interest, it would be quite hard and time consuming even to read it and to understand it. So, the topic makes a huge difference.

For the university student from Singapore (M SNNS) reading was an important skill that was essential for migrants in the area where he lived to integrate although the Chinese businessman had the opposite point of view, claiming that:

> You need huge vocabulary for the reading because those same words do not show up the second time, even some words you don't use in real life. You probably see them only in IELTS books…In the reading, you know, the topic is quite unique. No-one talked about that in the real life (Businessman, M CNNS).

This particular participant had been working in logistics for ten years so was used to using technical language in his daily work and yet he found no correlation between IELTS reading texts and the language he used on a daily basis for business purposes.

The reaction of the New Zealand employers I interviewed was that the ability to read well in English was an important skill to be tested. 80% of the employers noted that reading was important for

health and safety reasons, or to be able to read contracts and induction materials. Reading was most important for the business employer who worked in immigration because employees needed to understand government documents. However, even this employer stated that he thought the IELTS reading module was:

> Testing proficiency almost in a non-real-world environment. I mean I'm just looking at the stuff you're throwing across in front of me and it's almost like it's, you know there's a level of complexity here that perhaps isn't quite real world (Business Emp 7).

As for the educators, reading recent research articles was considered important in order to keep teachers up to date with innovations in teaching. However, even the university employer (Emp 1) and senior educator (Emp 2) were surprised at the topics and the lexical density in the IELTS reading section and questioned the reading age it was aimed at. The senior educator had been involved in literacy programmes for many years. Her reaction was as follows:

> Gosh, it's quite challenging. There's a lot of cross checking. I'd say that, given the literacy work I've done in schools, and even with quite senior students, lots of this would actually be quite tough. For people to actually learn just to focus because it requires not only to listen, read, speak and write, but also just to unpack what those things require from you cognitively because some of this is about reading between the lines and making deductions. It's going back and cross-checking what you've read with other bits of information so there's quite a lot of literacy skill required, and in a foreign language. Um, so I could actually say, hand on my heart, that some of this is not what you'd need to use averagely (Senior educator, Emp 2).

Other employers felt that both the reading and writing tasks were testing test-taking skills over English-language competence. The topics and time restraints were mentioned multiple times. The IT employer (Emp 5) noted that it would be challenging in real life because:

> So, you're reading and going back and forth, and you have an hour for this. It is quite technical, isn't it? That's interesting. You know this is actually harder than I had expected it to be. You know, there's a lot of vocabulary in this that…you could grab the average man off the street, and they would struggle with it, eh? Interesting. So, you need to go back and find the passage? I think that would be quite hard for anybody really (IT Emp 5).

The health and safety consultant (Emp 10) queried the topics and lexical depth in the reading section:

> The level of English is quite high. Um, some of our managers and supervisors could do that but not the operators, certainly not the operators. No, some of them couldn't go anywhere near that and they wouldn't do that in that amount of time…. No, these topics, they're not useful. Actually, what is ethnography?

Whereas the perceptions of the usefulness of the speaking and listening sections were generally favourable and they were partially favourable towards the reading section, the perceptions of all participants towards the writing section was more negative. It may be assumed that writing is a necessary workplace skill, and yet the role of writing in the workplace today seems to be changing (this will be discussed in more depth in the chapter 8). In general, the native speaker test takers wrote quickly and succinctly when given the writing topics. However, they found the writing section difficult for three reasons: the topics, time pressure, and because it came at the end of the test, at a point when they were quite tired. Apart from the university lecturer (M1) and the ESOL teacher (M4), none of the participants had needed to describe a graph in a written format in their places of study or work. They had interpreted graphs when reading information for work or study purposes but not actively written about them in descriptive terms. As the university lecturer (M1) pointed out:

> If they don't understand pie graphs maybe they wouldn't be able to write anything but that is not because they cannot write but it is because they don't understand the particular task. But, I think that people would struggle to write 150 or 250 words if it is not something that they have practised and prepared to do.

On the other hand, the managers who worked in offices found the topics for an argumentative essay strange because they had never needed to write on these types of topics for work:

> I think that the writing needs to be relevant, you know the writing for me it had no relevance for me whatsoever and the topics were completely…I just wasn't into it, you know… I think they need to make it more relevant to day to day life really. I think that would have helped (ICT manager, F2).

All of the participants remarked that for work or study purposes they had time to think before writing and that computers helped with the mechanics of writing such as spelling and punctuation. "But when I work on my computer, I do paragraphs, abbreviations, full stops and you have spell check which helps because not everyone is 100% on their spelling. So, I feel that if people were able to do the test on a computer system it would be easier" (Retail assistant, F3). The time pressurised environment of an exam left no time for proof-reading as well, although it seems that this is a common practice in workplaces. The ESOL teacher felt he had effective writing skills and had worked

for both private and public sectors, but as he pointed out “Of course, I had more time than in the exam so I could polish the skills” (M4).

In general, the non-native speakers accepted the writing tasks without questioning their validity. However, in a similar vein to the native speaker test takers, the non-native speaker participants also rebutted the idea that the writing section was connected to workplace competence. Nine out of ten of these participants felt the writing section of IELTS was unreasonable for the following reasons: lack of knowledge about the topics, time pressure, no time for proof reading, and the writing section coming at the end of a three to four-hour test. The primary school teacher (F GNNS) summed this up:

> My situation was that you have to wait around all day and then when you get in there, you’re locked in this room with 75 other people for three hours, more than three hours. It’s a really long time to have to sit somewhere, not being allowed to have a snack. You can stretch your legs by going to the toilet, but I felt like, you know, hot and your cheeks go red because you’re concentrating for so long. And it’s a lot of money, so you’re trying really hard, you’re concentrating really hard, and then at the end you have to write a 250-word essay on a topic that you have never even thought about before. And you have to write it and then you run out of time to double check it because the essay before that, the 150-word essay, was about describing a table graph that for me made no sense. So, I was sitting there thinking aah, and then the clock is ticking and if you are anything like a normal person you start feeling the pressure (F GNNS).

The writing topics for task two are described by IELTS as being related to issues of general concern. However, the general interest topics included in the IELTS test were not always perceived as relevant by the participants in this research. For example, the scientist from the Philippines stated:

> We don’t study topics like global warming. We read about it in the news. Like, I don’t have much to say about it and I’m not interested in it and I don’t know what would be the relevance when it comes, like, to the workplace (F FNNS).

Several participants mentioned that practising the topics in preparation classes helped them in the actual test but that this was not related to their writing skills at work. The Chinese businessman viewed the topics in the writing section as being a matter of ‘luck’:

> You know after the IELTS test we always check and ask, 'What's your topic?' and if you choose a really easy topic like pollution or something, like we did practise many times, we think oh lucky to get that (M CNNS).

Any misunderstanding about the writing topic could have affected the score results for the Filipino builder and carpenter because he stated that "Usually my writing is good, so I was expecting a good grade, but I was lost with the topic (M FNNS).

The university lecturer viewed the writing tasks as being 'doable', but time constraints made the task inauthentic "I would have loved to read through it because obviously when you are writing very quickly you are prone to making mistakes. So, I would have liked to review what I have written but there wasn't any time for that" (M GNNS).

Looking at productive output from an employer's point of view, it became apparent through analysis of the interviews that writing is not of paramount importance in the workplace today. It was interesting that the IT employer (Emp 5) said that she had good writing skills but "When do you ever write this sort of stuff? In general, in life nobody would have any idea nobody would have any idea what my written English is like". Surprisingly, even the educators did not value writing as much as oral communication for today's schools, commenting that writing was mainly confined to report writing:

> Certainly, the written communication skills, I would say, it's fit for purpose. So, it's able to communicate information to parents on the progress of their students, of their particular child. That's how I would view that … it's relatively limited report writing (University Emp 1).

The writing section came in for particular scrutiny from employers due to the time limit, the difficulty of the tasks and the task types being unrelated to any writing that they normally needed to undertake. The university employer (Emp 1), for example, commented:

> Oh, it's much harder than I anticipated, especially the writing part...I was always very capable with English and enjoyed it, but I'd struggle with this, straight up. I could describe those sorts of things … but excuse me, in twenty minutes, I would have to think about that. And, because of that time-boundedness, that's going to be really challenging.

It has to be borne in mind that this particular participant was a university researcher and teacher with a doctorate. Likewise, the IT employer (Emp 5) and business consultant (Emp 6) both felt that there was no need to communicate such complex ideas in writing in general at work:

> When do you ever write this sort of stuff? ... I would honestly find this totally irrelevant. I find this very odd. I'm actually quite surprised. It's not what I expected at all. I rather assumed there would be reading comprehension but it's the writing bit that I'm surprised by – that's a nonsense to me (IT Emp 5).

Regarding proof-reading, all employers noted that with computers, spell and grammar checks were used in workplace correspondence, so these helped remediate some syntactical errors. In addition, the primary school teacher (Emp 3) and business employer (Emp 7) explained that every item of correspondence, including their own, that was made public was proof-read before being sent out. This was because "There are bound to be errors and yet the standards for teachers and principals is high as a society" (Primary principal, Emp 3). Indeed, the IT employer (Emp 5) commented that her children's teachers' written work often contained errors, indicating that native speaker teachers do not necessarily produce 'perfect' language.

4.2.4 Receptive and productive written skills summary

While there was partial support for the reading module, the topics and lexical density of the reading passages caused some surprise for the native speaker participants. Many also commented that the specificity of the answers did not assess overall comprehension while general topics related to New Zealand life would have been preferred. The writing section was overwhelmingly seen as not being related to workplace competence. This was partly because of the topics, but also the lack of time for proof-reading. It appeared that most native speaker participants rely on computers or colleagues to ensure their writing is of a good standard.

4.3 Stakeholders' perceptions of the type of English being assessed

This section looks at what the research participants believed this language test was assessing after viewing, or taking, a sample test

As regards the test purpose, the native speaker participants felt that the test was assessing academic language proficiency in each section as opposed to language needed for living and working in New Zealand. The primary school teacher (F1) argued that:

> It's not assessing every day…it's assessing academic. It had a very academic bent which is why I asked you before who makes these things up. But, it was the …this doesn't assess the type of language you need to be able to be a competent member of society and to be able to interact with other people in this country...It's not inclusive because this is not English competency, this is academic competency and that is different.

For the native speaker test-takers who were not working in education, the test was declared to be difficult for people like them because they were not academic. “It was just too much information for me because I’m not academic” (F2).

Many commented that there should be different band score requirements for different professions and that a sheep shearer, a builder or a lawyer should not be assessed in the same way. As the retail assistant (F3) noted “Unless you were going to be a lawyer, then maybe that test is relevant but for someone who is more hands-on, practical, then it’s OTT”. On top of the professional level, it was noted that there are many types of English which are used in today’s workplace. Many native speakers mentioned that construction workers or IT professionals used their own types of language, not what was labelled as ‘Shakespearian prose’ or old-fashioned essays.

> The industry that I’m in, there’s a universal language and tech talk that gets through and makes them good at what they do but if you take one of my team, he’d just absolutely fall apart. He just wouldn’t manage this (the test) but he’s like a top-notch member of my team and he’s awesome (ICT manager, F2).

The variety of inner circle and expanding circle Englishes was mentioned with the ESOL teacher (M4) remarking that when he worked in Scotland, it took him three months to be able to understand his boss while, according to the university lecturer, in some workplaces in New Zealand it may be appropriate to understand Irish or Filipino English.

For the university lecturer (M1) effective communication in the workplace was essential. However, he did not feel that this type of test was assessing either workplace or academic communication since even for tertiary studies, communication as opposed to accuracy was paramount:

> I don’t think necessarily that spelling is a critical skill. And I think from a student’s perspective, in some cases we are marking for – if I knew what they were trying to say, even though it didn’t read very well, I would give them the marks because I wasn’t there to mark their prose… I was there to mark their understanding of economics (M1).

Although the university student (M5) could relate to some aspects of the test and was adamant he would have done better, had he studied for the test, he also found it annoying that he should have to prepare for what is purportedly a test of being able to communicate in English, his first language:

> I feel that annoys me because it shouldn’t require studying. It shouldn’t be testing how well you can study for a test of English (M5).

One of the most illuminating findings from interviewing the non-native speakers about their language testing experiences was that all of them recognised this type of test is a more of a reflection of educational level as opposed to communicative language competence. The reason why this is surprising is because, as a teacher, I had not reflected on this distinction, nor had it been discussed amongst my fellow teacher or fellow examiners. However, the non-native speaker test-takers saw the difference between basic interpersonal communicative strategies (BICS) and cognitive academic language proficiency (CALP) as being especially pronounced in the reading and writing modules. For example, the Filipino scientist recognised that although it is called an English test, "It must be a test of education levels, not communication (F FNNS)." The ESOL teacher could see the values of the writing module for post-graduate study but felt:

> I think that unless you are in the academic field or that you are constantly report writing, then yes, but if you are migrating and you want to work in, I don't know. If you are perhaps a nurse, I don't see the relevance. I think written...Do you really need to be able to write to this extent for migration purposes? That's my question (F MNNS).

They all expressed doubt as to whether people working in manual jobs should need to take a test of this nature. While these participants noted that the reading module was not beneficial to everyday life, the writing module attracted the most comments; this could be attributed to the fact that 90% of the non-native speaker participants were disappointed with their writing score results. The businessman from China argued that IELTS did not test skills which were useful for everyday life. He had been relieved to discover that everyone he worked with in a range of companies understood his writing and due to his workplace experience he had been accepted on a Master's in Business Logistics, even though he had scored only 6.0 in IELTS. Age was a factor that was mentioned on top of educational level by the Filipino builder:

> I would say it tests education not communication because when we talk about the communication, I believe most of the Filipinos do understand what is happening but when it comes to education, there are lots of Filipinos who fail. I would say, in my level, because I am now 42 years old, I believe that in my younger age I can pass the IELTS (M FNNS).

For the Vietnamese retail assistant, this type of test was not only for well educated people but for those who were proficient users of the language. She claimed that native speakers would only get over 6.0 if they were well-educated (F VNNS) and that the results would be similar for Vietnamese taking this type of test in their mother tongue.

> In my opinion, for example in Vietnamese, just experts in Vietnamese can reach 9.0. So, I think to get a high score you need language that is very exact. I mean you find luxury in that language (F VNNS).

This opinion was echoed by the businessman from China who likened IELTS to reading and writing Chinese poetry "IELTS is exactly just like that, like using the old Chinese when you have to write special things" (M CNNS).

Bodies such as the Medical Council or the Teaching Council of Aotearoa New Zealand | Matatū Aotearoa accept only the Academic version of the IELTS test. However, the employers in the education sector thought a language test for migration purposes would be assessing general communicative competence. For this reason, they queried why the types of tasks they viewed in the sample IELTS test were being used as a measure of functional English. The primary school principal had recently started studying for a Master of Business Administration and admitted that he had needed to adjust his style of writing from what he used in his workplace. For him, a high level in IELTS did not translate to workplace competence but rather academic proficiency:

> So, a level 7.0 for me is meaning that they're more academic. They can write in a more academic style, which they may never do. From the time they get a job, they have a successful life, wonderful citizens here and never again might they have to write like that, so what's the purpose? (Emp 3).

He concluded that primary school teachers do not need to write academic essays, and for this reason, he did not perceive this type of language test as being fair or relevant for people coming to New Zealand for immigration purposes.

The situation may be worse for refugees and migrants. The ESOL employer (Emp 4) worked with adult students who needed to take IELTS for immigration/work purposes. She had been under the impression that IELTS was supposed to be testing general English and pointed out that this type of test was impossible for some of her students because some migrants, particularly women, had barely been to primary school. For her, even those with a good level of education do not need high test scores "Others are quite well educated but even then, they're not going to be needing this level of English. Maybe I'm wrong about 6.0, but it's quite high – maybe a 5.0 but it's still quite a lot" (Emp 4). When she viewed the IELTs examiner's remarks (publicly available material) which mentioned that the test-taker had included irrelevant material, she remarked that the test must be assessing test-taking skills rather than English ability. She inferred from this that "The better educated you are

in your country, the better you'd be able to perform on that because you'd understand exam questions… you'd understand how to perform in exam conditions." (Emp 4).

Knowledge that IELTS was initially utilised for entrance to tertiary education led some employers to point out that, "I think it's measuring proficiency according to, and it's a bit subjective, but according to something that was never made to be a workplace measure" (Business consultant, Emp 6). The business employer (Emp 7) had been involved in immigration since IELTS was first used. He was of the strong opinion that English was crucial for migrants working in New Zealand, but English for the real world rather than for this type of test which he regarded as being English for academia. The retail/manufacturing employers commented that the type of English being tested in the samples they saw was not relevant for their workplaces at all and queried what sort of education level was needed to take the test because even they were unable to understand some of the vocabulary being used.

4.3.1 Summary

All of the participants in this study, native and non-native speakers alike, felt that this type of language test is more of a reflection of educational level as opposed to communicative language competence. For this reason, it was not deemed suitable for low to semi-skilled workers.

4.4 Stakeholders' perceptions of how well they can use and interpret test score results

This section includes test takers' (both native and non-native speakers) perceptions of their own, or their likely, test score results as exemplified in tables 6 and 7. This is complemented by the test users (employers) perceptions of the test score results of the native and non-native speaker test candidates in order to explore their feelings about the required minimum cut-off scores to live and work in New Zealand. These data provided evidence for RQ2: How well do test takers and test users feel they can interpret and use test score results?

4.4.1 Native speaker test candidates' perceptions of the test score results and their meaning

After finishing the test, the native speaker candidates could view their test results for the reading and listening sections, but they were not given scores for the writing section for ethical reasons. This is because the two examiners had been IELTS assessors for 15+ years and had signed an IELTS agreement not to discuss possible test scores with test takers. Instead, the native speaker test takers were given sample answers written by non-native speakers of English to comment on regarding the level of English they believed the writers of these samples had, and whether the writing would be

acceptable in their workplaces. Regarding their own results, table 6 shows that the highest scores were on the listening section with three participants scoring 6.5 and nine above this threshold. At present a score of 6.5 is required for permanent residence in New Zealand while a score of 7.0 in all sections is required by The Teaching Council of Aotearoa New Zealand | Matatū Aotearoa for teacher registration, the Immigration Advisors' Council and the Medical Sciences Council (Three of the non-native speakers in my research had registered with these three councils). As for the reading section, five participants scored 6.5 or under, with seven scoring above. Interestingly, the two participants with the highest score (8.0) were amongst the youngest (in their early twenties), which belies the statements made by the older participants that young people in New Zealand would struggle badly with this type of test.

The section with the largest anomalies was the writing section where eleven of the native speaker test takers (91.6%) scored 6.5 or under (task one) and seven (58.3%) scored 6.5 or under for task two. It is important to note that, although this was not an official exam, seven native speakers scored 5.5 or under for task one and five scored 5.5 or under for task two. Such scores are lower than those needed by non-native speakers of English for immigration, employment and professional registration in the New Zealand context. The question, also raised by participants in this research, is why migrants to New Zealand should be expected to gain higher scores in the writing section than the average New Zealander could realistically achieve. For the Teaching Council of Aotearoa New Zealand | Matatū Aotearoa, these results highlight that the written literacy skills of native speaker teachers and lecturers, together with other professionals (such as those in the IT sector) that they may want to attract to the teaching profession may be lower than those of migrants applying for registration.

Table 6. IELTS test results for native speakers of English

Females

	Age	Listening	Reading	Writing Task 1	Task 2
F1	60s	8.5	7.0	7.5	8.0
F2	50s	6.5	6.5	5.5	5.0
F3	40s	6.5	6.0	6.0	6.5
F4	20s	8.5	8.0	6.0	7.5
F5	40s	8.5	7.0	5.0	5.5
F6	30s	8.5	7.0	5.0	6.5

Occupational areas

F1 – Primary school teacher

F2 – ICT manager

F3 - Retail assistant

F4 – Waitress

F5 – Accounts manager

F6 – Master's level student

Males

	Age	Listening	Reading	Writing Task 1	Task 2
M1	40s	8.5	7.5	6.5	8.0
M2	60s	6.5	6.0	4.0	5.0
M3	20s	7.5	6.0	5.5	4.5
M4	60s	7.5	7.0	5.5	7.5
M5	20s	8.5	8.0	5.0	7.0
M6	teen	7.0	6.5	5.5	4.5

Occupational areas

M1 – University lecturer

M2 – Telecom technician

M3 – Supermarket worker

M4 – IELTS/EFL teacher

M5 – First year undergraduate

M6 – Year 12 student (high school)

4.4.2 Non-native speakers' perceptions of the test score results and their meaning

The non-native speaker test-takers felt that the test was reliable, mainly because it was an international test taken by millions worldwide. They all felt they were aware of what different band score results signified and, unlike the native speakers of English, they also recognised that there was a big difference between half points (for example between a 6.0 and a 6.5).

Since the majority of the non-native test-takers had taken the IELTS test more than once and had attended preparatory classes or viewed IELTS test-taking skills on-line, they claimed they knew what test assessors were looking for and how to improve their scores through test-taking strategies. The university lecturer (native speaker test-taker, M1) voiced concern about this, referring to cases he knew about where prospective university students could do well by learning these test-taking strategies but not be able to communicate in English. Indeed, such strategies were mentioned by 80% of the non-native speaker test takers "I just focus on how to pass it, not how to understand it." (M CNNS). The builder from the Philippines outlined many of the strategies he had been taught for the four skills in the following extract:

> We had strategies. For example, in the reading how to get the answers and not to read all the story, all the sentences. And, there are techniques like finding the key words. In speaking, they try to teach us how to be confident and how to deliver our sentences. We always have these strategies in listening like how to expect the answers because you need to read the questions first before you listen, so we need to predict the answers, whether it's a number, date or names. And for the writing, they gave us the proper format, which is how to deliver writing in formal and informal ways. Yes, and which comes first and what is the body and how do you address your letter. And for task two, how can you explain the topic by topic sentences and then elaborate and give your idea, your view (M FNNS).

Other non-native speaker test-takers mentioned knowing certain "tricks" that are expected by those marking the test. In particular, they described words that they had learnt to include such as 'frequently' (F MNNS), and 'first, secondly and last but not least' (M CNNS). However, as the librarian from Poland and the business man from China explained, these were words that they never used in their normal correspondence at work:

> It's valued if you don't use too many words the same, so you are not repetitive or if you use just sophisticated words like moreover, nevertheless or ultimately and that gives you some points regardless ... whereas Kiwis don't use them at all, unless they are politicians (F PNNS).

The antithesis to this is evident with the Vietnamese retail assistant who scored well in the writing module with a 6.5 but felt unable to communicate satisfactorily in every-day life while the businessman from China knew many compatriots who had learned 'how to pass' IELTS but could not converse in everyday situations in New Zealand. He illustrated this by giving the example of Chinese students who had scored well in the IELTS test but did not have sufficient communicative English to top up their phone cards.

The notion of 'luck' impacting test score results was remarked upon with the businessman explaining that candidates were fortunate if the reading and writing topics were similar to those practised in a preparation class, but "if you get a topic that people don't usually talk about, we think 'oh no bad luck', which means it doesn't really test your English skills, it's all about technical (technique) and templates" (M CNNS). This was reiterated by the production manager (M INNS) "So with reading and listening it depends upon your luck" because, as he recognised, missing just one or two questions makes a band difference. The university student from Singapore regarded English to be his first language so when he received 7.0 for his writing his reaction was "I don't think there is something wrong with me. I just think maybe bad luck with the person who is marking it".

Table 7. IELTS test results for non-native speakers of English (range of results as reported by interviewees)

Females

	Speaking	Listening	Reading	Writing
F GNNS	9.0 8.5	7.5 7.0	8.5 8.5	6.5 6.5
F MNNS	8.0 – 9.0	8.0 – 9.0	8.0 – 9.0	7.0
F PNNS	7.5 – 8.0	7.5 – 8.0	7.5 – 8.0	5.5 – 6.5
F FNNS	8.5	8.0	7.5	7.0
F VNNS	6.0	4.5	6.0	6.5

Occupational areas

F GNNS - Primary school teacher (German)

F MNNS - ESOL teacher and PhD student (Malaysian)

F PNNS – Librarian (Polish)

F FNNS – Scientist (Filipino)

F VNNS - Retail assistant (Vietnamese)

Males

	Speaking	Listening	Reading	Writing
M GNNS	8.0 – 8.5	8.0 – 8.5	8.0 – 8.5	6.5
M SNNS	8.0 - 9.0	7.5 – 8.0	7.5 – 8.0	7.0
M CNNS	6.0	6.0	6.0	6.0
M INNS	7.0	7.0	6.0	6.0
M FNNS	6.0	5.5	5.5	4.5

Occupational areas

M GNNS - University lecturer (German)

M SNNS - University student (Singaporean)

M CNNS – Businessman – logistics (Chinese)

M INNS - Production manager (Indian)

M FNNS – Manufacturing worker (Filipino)

4.4.3 Employers' perceptions of the test score results and their meaning

The group who trusted test score results the least were the New Zealand employers. The IT (Emp 5), ESOL (Emp 4) and immigration (Emp 7) employers indicated that they had very little awareness of the meaning of scores. Confusion arose because potential staff members regularly presented both these employers with IELTS test results when applying for jobs but the IT employer (Emp 5) and immigration business employer (Emp 7) found no correlation between their scores and language competence in the workplace:

> I don't take any notice of it...I haven't noticed any great correlation between IELTS certification or the level they have passed and the standard of spoken English and written English, so we tend to make our own assessments and judgements (Emp 5).

> As an employer, if you slapped down an IELTS and said I got 8.0 in all of them, that wouldn't be enough to satisfy me because I wouldn't see the test as that important. And, if they got 6.0s in everything, I wouldn't see that as important either (Emp 7).

Similarly, the ESOL employer (Emp 4) never asked for an IELTS score because "I don't think there's any relevance. I don't know why. It must be the test or maybe it's the complexity of the workplace and it can't be summed up in a test".

The business employer (Emp 7) mentioned he was confused when he compared the IELTS scores with educational achievement because some of his employees had received a Master's level degree in New Zealand but could not gain 7.0 IELTS. Conversely, this employer (Emp 7) had met people who had attained 7.0 IELTS, and yet, they could not communicate orally. Since the majority of the employers in this research valued oral communication over written communication, this may again indicate that a high score in the speaking section is most important for social integration and workplace competence. Insufficient connection between test scores and workplace competence led the IT employer (Emp 5) to claim:

> You know, we have had people who have apparently had very high scores on this…but I haven't thought 'wow' your English is awesome ... As I say, I wouldn't see these as a strong predictor of someone's ability to be a valuable employee (Emp 5).

The university employer (Emp 1) had noticed that if prospective non-native students for the Graduate Diploma in Education had problems enrolling due to the language pre-requisites, this was because their writing scores were lower than their scores for the other three skills, but he was at a loss to understand why this was the case. It is interesting to note that the senior educator (Emp 2) believed candidates should gain similar score across the four skills unless they had learning difficulties in one area whereas the business consultant was aware that "it's very rare that you get someone who is amazing across all four skills" (Emp 6). Initially, the senior educator (Emp 2) was under the impression that an IELTS 7.0 meant a person was "able to operate at a minimal level in language proficiency" (Emp 2). However, after viewing a test and some test responses, she concluded that "Well, 7.0 is not really minimal. It's actually pretty high" (Emp 2).

4.4.4 Interpretation of score results

As for interpreting the test score results, the numbers that are given on a test score report form could not be deciphered by the native speaker participants (both native speaker test takers and employers).

Regarding their own test score results, most of the native speaker participants were surprised by their scores. The economics lecturer and the ESOL teacher (M4) thought they should have scored higher while the retail assistant (F3) commented that:

> Well realistically as a native speaker I should be able to pass it with flying colours because I'm a New Zealander and I live in New Zealand and I speak English every day so I should be

> getting high scores, but I didn't. I think we should be scoring really high, like 7.5 or 8.0 but we didn't.

The two participants who worked with international students should theoretically be aware of what test score results mean and how they relate to performance in the real world. However, they found it particularly difficult to interpret the test scores in writing section, and the test scores were not seen as representing useful or useable skills for today's workplace. The ESOL teacher (M4) could not understand why a student he taught was struggling to gain more than 6.5 in the writing section when she was achieving 7.5 overall. He indicated that "As far as I'm concerned, she would go away, and she would write these jolly wonderful essays and I think she wrote them a lot better than I did." The university lecturer (M1) test taker had knowledge of the score results of his own students but could see no correlation between these results and their academic achievement "I'd actually always wondered if there was a difference between a 6.5 and a 7.0 and if this actually meant anything" (M1). This lecturer's lack of comprehension of the meaning of test score results also came from the fact that he had observed that many non-native students who achieved the prerequisite scores for university entrance could not necessarily communicate well in English.

Interestingly, the task which describes a graph was identified by the IT employer (Emp 5) and the ESOL teacher (M4) as being useful for a business presentation:

> Task 1 Academic is for business writing. It is for business reporting. You don't have the luxury of presenting this in whole, long complicated sentences because this type of writing will only ever be used if you're sitting in a business meeting, you've got an agenda with 40 items on it, you've got the board meeting around you and you've got two minutes to present what you're supposed to say so your sentences need to be short, succinct and to the point because that's the nature of business writing. And if they're marking this in the irrelevant standard, they are doing the students a grave disservice (ESOL teacher, M4).

This statement highlights that even the ESOL teacher was not aware that both IELTS writing tasks are assessed on the breadth and depth of morphosyntactical and lexical language presented by candidates. When commenting on how the written sections are graded, the ESOL teacher explained that for him, it was "as clear as mud" (ESOL teachers, M4).

To evaluate which test score results in the writing section would be acceptable for the New Zealand workplace, the native speaker participants (test takers and test users) were shown some sample responses written by non-native speakers of English which were graded as being between 4.5 and 6.0 in publicly available material. They all remarked that this level of writing was acceptable for the

New Zealand workplace. Surprisingly, even a response which had scored 4.5 was regarded as being adequate "And that's a low fail but that seems quite articulate to me" (Retail assistant, F3). They tried to decipher why there were any issues with the writing:

> They are very well written, very easy to read, very clear. I think, it's kind of like what I was trying to say... So, what lost them the points? What have they done wrong? (University student, M5).

Moreover, they could not understand how the writing had been assessed because the scripts appeared to be as good as or better than their own writing. Typical comments were:

> "But that's ridiculous because that is really good. That's crazy, really crazy. If I could write like that, I'd be really happy. That's so wrong because that explains it quite well. So, who marked them on that? Is it just because they haven't got their punctuation? (Telecom technician, M2).

For the task one response which had been graded as a 6.0, the Master's level student felt:

> This one is on the carbon dioxide. Apart from one or two spelling errors, this one is academic, and I think it would score much higher than mine. I'd definitely give it a 7.0 plus, I'd say (Master's level student, F6).

In general, the non-native speaker test takers were used to taking IELTS and accepted the test scores as being reliable. The only section they queried was again the writing section. The university student from Singapore was confused because, while he recognised that high stakes tests are looking for what he called a "cookie-cutter standard", an essay he had submitted to university scored an 'A' grade, but he received only a 7.0 for a similar essay he had written for the IELTS test. This scenario was the same for the primary school teacher from Germany who had studied literacy at university for three years and had scored A+ in assignments but the highest she had scored in IELTS for her writing was 6.5 (she took the test twice). The Singaporean student's summation of this anomaly was that the criteria for the two sets of writing must be different, stating that with IELTS "There is a set structure and if you deviate from that structure, they actually mark you down for that even though it might be acceptable elsewhere" (M SNNS). While some test-takers (both native and non-native speakers) felt that the marking must be subjective, the ESOL teacher from Malaysia had received training from an IELTS examiner and was aware that:

> It's very much a rubrics-based because it's so standardised and people can't mark on what everyone feels so it's 100% based on rubrics so if you can master the rubric you can get a better score (F MNNS).

Another issue raised by the research participants was the way in which the scores are reported, the limited amount of feedback given to test takers, and the lack of information about what a test score means in real world terms. A lack of feedback on the test scores caused consternation for a few participants as they had no idea how to improve their English for this test. With the scale of high stakes language testing, it is not possible to give individual feedback; rather, the test score result feedback is generic with every candidate who scored 6.0 on a given skills area, for example, reading, receiving the same feedback. This leads to a lack of *Opportunity-to-Learn* which is defined by Kunnan (2018) as being an important element of test fairness. In short, test candidates should be able to learn from their mistakes. However, as the primary school teacher remarked:

> You just get a general synopsis and out of that synopsis you have to pick what you did wrong, which I didn't. I couldn't really make much sense of it, to be honest. I was quite shocked (F GNNS).

The university student (M SNNS) echoed this frustration, explaining that at university he received feedback from lecturers but with IELTS "It's quite difficult to say because it's quite isolated and we don't get the papers back. Sometimes I would like to know exactly what I did wrong" (M SNNS).

The band score results may have some meaning for test-takers who need to sit IELTS numerous times but the university lecturer from Germany (M GNNS) argued "there is no meaning to numbers" for decision makers who often have very little idea of how to interpret these scores. The ESOL teacher from Malaysia suggested that there were wider implications because "If the highest score is a 9.0 and you get a 7.0, people always think what's happening to the other two points? Are you good enough to write reports and things like that?" (F MNNS). While decision makers may regard band-score results as representing a stable trait which identifies the English level of a test taker once and for all, none of the non-native speaker test-takers who had taken the test more than once had received exactly the same score. Instead they reported the range of scores achieved (Table 6).

The employers were unable to interpret the test score results in any section. The results for the speaking listening and reading sections were numbers for them devoid of any real meaning. To help them understand the writing section they were shown sample test responses for this module written by native speakers of English from Phase One of this research (the native speaker test takers) and non-native speakers of English responses (published material). There were two sample responses for

Task One (one native and one non-native) and two Task Two sample responses (the grades given were between 5.5 and 6.5). The employers were asked to comment on the standard of English, and whether they could identify which ones were written by native or non-native speakers of English. All of the employers regarded the four sample scripts as being adequate for their workplaces. The majority of these employers queried if the test was assessing test takers' English or their knowledge of the rubrics, and whether a score of 6.5 or 7.0 was appropriate for writing in today's workplace. The primary school principal (Emp 3) was doubtful of the relevance of the scores for the writing section:

> Can you be understood in what you're trying to say in your writing. Now, all of those four that I looked at, … they got their point across. You know, there was informal language used in parts and a spelling mistake in one, but the message was clear. As far as I'm concerned, all those would be fine for a primary school setting (Emp 3).

On top of finding all four scripts acceptable for the workplace, the employers struggled to identify which had been written by a native or non-native speaker. Although they were not asked to give an estimate of the band score results, a few did. The senior educator was typical since she assessed all the scripts (below) as being higher that they had actually scored.

Table 8. Senior educator's predicted scores for scripts

Senior educator	**Predicted scores for scripts**	**Actual scores**
M1 task 1	8.5	6.5
NNS task 1	6.5 – 7.0	6.0
F5 task 2	7.0	5.5
NNS task 2	6.0	5.5

It is not possible to give a conclusive summary of the employers' remarks in relation to the merits of each of the sample scripts because many preferred different sample responses. However, it is surprising that the majority selected the response to the task one graph written by a non-native speaker (who scored 6.0) as being the best. The second most preferred script was written by the accounts manager (F5, who scored 5.5) while only one chose the one written by the university lecturer (M1, who scored 6.5). The only consistency was that the essay written by the non-native speaker (who scored 5.5) for task 2 was regarded by some employers as being of a lower standard, but perfectly acceptable for the workplace.

When explained that all four written samples had scored below the requirements to be a teacher in New Zealand (7.0), the primary principal (Emp 3) responded “Gee whizz – so what are they lacking?” while all the business, IT, retail and manufacturing employers had similar responses to the business employer who declared: “Well, I’m surprised at how low some of those scores were” (Emp 7) and “I would say in the workplace that is pretty high” (Emp 6). In fact, the retail and manufacturing employers regarded all four sample test responses as being of a much higher standard than any writing they generally read in their workplaces, whether written by native or non-native speakers of English.

4.4.5 Summary

Non-native speaker test takers learned test taking strategies but recognised these strategies were not useful for living/working in New Zealand. The non-native speakers trusted the test score results in the speaking, listening and reading sections, but they felt the scores they had received in the writing section were not comparable with their writing competence for work or study. The native speaker test takers did not score highly in the writing section with many receiving approximately two bands lower than for the other modules. Surprisingly, even the ESOL/IELTS teacher (M4) did not understand the grading for the writing section. Employers could not comprehend the meaning of test score results in any way. They could not see any correlation between test scores and employees’ or students’ English language abilities.

CHAPTER 5

FINDINGS PART TWO: PERCEPTIONS OF ENGLISH LANGUAGE ASSESSMENT AT A GROUP OR WORKPLACE LEVEL

5.0 Introduction

In this chapter the macro implications of using test score results are discussed. Specifically, the main question to be addressed was: How is language used in the contemporary workplace and how does this compare to test content? The participants' perceptions (native speaker and non-native speaker test takers and native speaker employers) provided empirical evidence for RQ3: How well do test takers and test users feel that the IELTS tasks and the constructs being tested reflect communicative language used in the New Zealand workplace? Further, all of the participants gave their views on the general literacy skills of native speakers of English in the New Zealand workplace.

5.1 How language is used in today's workplace - evidence for the extrapolation inference

Language is used in different ways in a variety of workplaces. However, to assess test validity and fairness, participants (native speaker and non-native speaker test takers and native speaker employers) in skilled (education/IT) and non-skilled (manufacturing/retail) fields were asked to share their perspectives on the type of linguistic skills which are valued in the modern New Zealand workplace. It transpired that the main tasks performed by employees included using templates, brief reports and emails, and all these tasks are nowadays aided by technology. The topics that were deemed important were those related to policies and procedures, and in particular health and safety regulations. According to all the participants (native speaker and non-native speaker test takers and native speaker employers), language use in the workplace has changed, and is still changing, with communicative competence and simple, clear English being valued over precise, syntactically and lexically complex language.

5.1.1 Authentic task types

5.1.2 Filling in templates and forms

When the majority of my participants (native speaker and non-native speaker test takers and native speaker employers) compared workplace tasks with the sample language test tasks they were given, they commented that the ability to write an essay is not generally beneficial for the workplace. Many native speaker test takers did not see the importance of a coherent and cohesive piece of writing because the amount of written output produced in most workplaces is short and often facilitated by templates and forms. The native speaker test takers in the IT and manufacturing fields explained that:

> You sometimes have to write documents but not everybody needs to. Filling in forms on-line and maybe brief explanations of this goes there or that goes there – pretty simple stuff really (Telecom technician, M2).

This corresponded with the work-related tasks outlined by the ICT manager (F2) "And we have standard templates and things like that. And just use standard things that you are just plonking in bits and pieces".

The amount of written English needed for retail and customer service was more limited as in the situation described here by the native speaker waitress (F4):

> Everything you learn is just verbally learnt. You don't, like, write anything down. You are shown how to do it and they tell you how to do it and that's it, you're done. You don't have to read or write anything for that ... maybe a small amount of writing, just so that they could just like write key words for the orders but not like a full-on essay. They'd never need anything like that for work.

The supermarket employee (M3) reiterated this, saying: "The only reading and writing we have is, say we're taking the chicken out of the oven and we have to temp it and we have to write down the time and the temp and the date ...That's literally it. Just writing in records".

This comment was echoed by the health and safety consultant (Emp 10), who remarked: "They only have to write down numbers like cooking temperature of 80 degrees. They don't have to write full sentences" (Emp 10), while in retail the situation seemed to be the same: "For the writing, we don't write much because it's your name, phone number and what you would like to order (Retail director, Emp 9).

Commenting on the types of tasks that migrants may need to complete, the telecom technician (M2) regarded simple applications as being important:

> In terms of someone coming into NZ, you know, all those basic skills…to write a CV, to do a job application, to do a … it's not about tricky questions to catch them out on finer shades of meaning. It's more about being able to communicate clearly but I think this goes beyond that. It's asking for a fairly high level of English ... but, if they need to write to that level is another story because you just don't typically. You don't have to write essays (M2).

Likewise, talking about her compatriots, the laboratory scientist from the Philippines pointed out that writing in an exam specific way is not a task that is useful for migrants who have come to New Zealand to work as tradespeople, and that writing is a skill which they rarely use:

> And writing all the more. Their job wouldn't entail them to write all these statistics and all these graphs. Why would that be a test of their capacity or their ability to work and why would that be a hindrance for them to migrate, just because they can't interpret statistics and write it in the way IELTS wants? (F FNNS).

5.1.3 Policies, processes and procedures

Apart from filling in forms and templates, from the perspectives of the native and non-native speaker participants in this study, it appears that reading and writing tasks in the New Zealand workplace are completed, in the main, to explain and comprehend policies and procedures. As the Master's student (F6) who also worked in a production company explained, all tasks followed processes, controls and procedures: "Um, process, control, procedure, so basically step by step instructions on how to do the job. And that's what everybody has to follow" (F6). Even when tasks involved more writing, the amount involved was not great since "For the workplace, you pretty much have to write some type of summary of what went wrong… you do need to write down maybe a small summary. A few words about what went wrong, what went right and how things could be done differently" (F6).

In terms of listening skills, the predominant ability that employers expected their employees to have was understanding processes through instructions: "It's actually pretty rare for someone to sit there and listen to somebody for ten minutes straight. Short, sharp instructions is what they need to be comprehending" (Emp 6, business consultant). This was corroborated by the IT employer (Emp 5) "For the technical roles, they just need to be able to understand and follow instructions and communicate technical ideas".

5.1.4 Health and safety

The ability to comprehend policies and procedures may be linked to the topic that was overwhelmingly described as being crucial for today's workplace: health and safety. The majority of participants (test takers and test users) mentioned health and safety. The manufacturing employer (Emp 8) encapsulated this:

> From a health and safety point of view our requirements are becoming much greater and it's not just about can people read the safety sign, it's can they comprehend what they need to do with that and it's how they need to respond in certain situations, so the comprehension is becoming more important to us. Health and safety is becoming such a big part.

For the business consultant (Emp 6), this did not necessitate high level literacy skills unless employees wanted to gain employment at a supervisory level. As she pointed out:

> In construction, it often is just the speaking and listening that's the most important. They need to understand those instructions, but for health and safety purposes ... They can't reach that next level up if their English isn't that good because health and safety is important at a supervisory level ... And for those lower skilled roles, they do not need any higher than that (IELTS 5.0) They are not writing onerous reports. They are not writing reports at all. They might have to fill in forms and often construction companies are pretty good at making sure that they understand those. Maybe, they are translated, for example. For health and safety, they do have to make sure that they are covered in that area.

While the retail employee (M3) commented that in his role, "I had to write about food safety and health and safety practices", the amount of written language needed for this purpose was limited. Rather, the waitress (F4), retail employee (Emp 3), ESOL employer (Emp 4) and business employer (Emp 7) indicated that health and safety regulations (for native and non-native speakers of English) were covered more through receptive skills (listening and reading) rather than in a written manner with comprehension being assessed orally:

> We have a lot of health and safety but that's also more verbally done through a presentation and we just go through it on a sheet and then the manager verbally goes through it with you and they tick it off to make sure that you know what they're going through… For health and safety, basically you need to be able to write one word or one sentence. You don't need to write a whole sentence on it (Waitress, F4).

> They may be in a job where they mainly need to be able to speak but they need to be able to read to understand the health and safety and other instructions because that's a big thing nowadays (ESOL employer, Emp 4).

> The carpenter doesn't need such a high level of English, but doesn't he need to be able to read plans an all about the health and safety issues? You know, doesn't he need to clearly understand that? And, in the workplace these days, having a good command of English, certainly at the spoken level is really important (Business employer, Emp 7).

The ability to respond to texts in a spontaneous way, as in a testing situation, was not considered to be productive for the retail employer (Emp 9). This participant explained that staff were given a couple of weeks to familiarise themselves with health and safety and other workplace procedures

because: “We give them a couple of weeks to make sure what they’ve got is clear and we don’t need someone who is too bright.”

5.2 The use of technology

5.2.1 Emails

Having worked in New Zealand for many years, I have used email on a daily basis, and therefore anticipated the importance of email being mentioned. However, my horizons were changed by the responses of my research participants. Emails were not mentioned as much as templates and health and safety. The employer who did mention emails was the business employer (Emp 7) who noted that a key part of their work as immigration advisors consisted of emails and written submissions to the government.

As for the non-native speaker test takers, the librarian from Poland (F PNNS) had just changed jobs and felt that she would now need to read many emails: “In this new position, it will be quite a lot (of emails). I will probably have to reply as well but not 250 words [as for Task 2 in the written module of IELTS], not at all”. The test taker from India (M INNS) was a production manager who predominantly needed to read emails as opposed to other types of reading. Therefore, for him IELTS reading tasks did not correspond to workplace tasks “The reading section, it’s not so helpful because in the reading side I only need to read emails and all that on the computer so it’s only 1 percent helpful in my workplace.”

However, this production manager (INNS) argued that the writing section of IELTS was beneficial for writing emails in the workplace because:

> In IELTS if you go for General, you need to write a letter, formal or informal, so if you have experience of formal/informal letter, you know very well how to write a good email to your general manager (M INNS).

Nonetheless, he explained that these emails were generally only one to two lines in length and that in his workplace “we don’t use that much reading and writing” (M INNS). The Chinese businessman reported that he wrote emails every day but, interestingly, the people he corresponded with understood even if he used inaccurate English or phrases in Chinese:

> Well every day, I probably reply like at least 10 emails with my custom brokers or MPI and they all understand what I am saying. They are pretty open, and in real life, you know, it depends how you deal with the people. I’ve been working with my custom broker for like

> three years and sometimes I put some Chinese words and he understands (laughing). …. They understand what I am saying. The grammatical errors don't matter (M CNNS).

5.2.2 Technical and work specific jargon

In general, in business and IT, the nature of the job tended to indicate which of the four language skills were relevant with reading and writing as opposed to oral skills being necessary, for instance, for code writers. Yet, the language they use is not necessarily English, as the business consultant (Emp 6) explained:

> You know I talk to some of my IT people and they're like, we don't need them to be excellent at speaking and listening because most of the time they're reading, writing, and then they have codes and a whole lot of other language.

It seems that the language used in the workplace, which may include codes, acronyms and jargon, is universal, crossing country borders. The ICT manager (F2) commented:

> You know, because of the industry that I'm in, there's a universal language and tech talk that gets through and makes them good at what they do … And like with my team, you know they send out emails and their language is techy. It's talking about Windows 10, it's talking about the task manager, it's talking about the things that people understand globally and when it comes to grammar, emails, they've got the spell check (F2).

Therefore, rather than precise, accurate English, most participants in this study felt that technical language and jargon need to be used and understood with the first language of staff members not always being so relevant. This was identified by the Chinese businessman (M CNNS) who worked in logistics:

> If I gave native speakers a quote and it's got the acronyms like B and L or B and O, or customs delivery order or what does BAAC stand for, they wouldn't understand even if they are native speakers so why are you judging people, their English skills, by IELTS?

The contrast between academic language and workplace language, where numbers and jargon seem to be common, was identified by the laboratory scientist and the librarian. The Filipino scientist (F FNNS) voiced this difference, stating "In the lab we deal with numbers. We do have reports, academic writings, we do, but I don't think we use all those fancy words". Similarly, the Polish librarian was of the opinion that people working in occupations as disparate as vets, nurses and chefs use a lot of jargon while kindergarten or primary school teachers need child appropriate language. She, therefore, questioned the relationship between IELTS and workplace language:

> So why do you need 7.0 if you go for a primary school teacher? Well, I think that was just a bit over the top. No, I think that was unfair ... With IELTS you need to use academic words to score 6.5 so if you are maybe a chef, where does that go – into the soup? (F PNNS).

While codes, jargon and client specific language may be more relevant for workplace competence than English, the need for migrants to understand New Zealand jargon was also mentioned. The business consultant (Emp 6) worked with a wide range of industries and felt that language needed in the workplace is "Simple but also a LOT (her emphasis) of jargon. And we assume that our newcomers understand all this Kiwi slang and industry specific jargon, but often they don't."

5.2.3 Technology – friend or foe?

Regarding the role of technology and how helpful it may be for English language accuracy, the participants' remarks did not prove to be conclusive with some pointing out that computer editing tools were helpful while others felt they hindered a user's language ability. The manufacturing worker from the Philippines spoke of writing by hand as being old-fashioned: "But as I said earlier, we are not using writing in our jobs and if we do writing, we just send them email through our laptops. We don't usually do the old-fashioned way of writing" (M FNNS). Many native speakers noted that they only wrote with the help of computers and that these helped with proof-reading, thereby circumventing the need to know the rules of English grammar and spelling. The native speakers who mentioned the advantages of computer editing were the accounts manager (F5) and retail assistant (F3). They valued technology because:

> It's more computer-based skills in my workplace so spelling isn't important because you have spell check. They have to be able to interpret data and probably put it in bullet points (Accounts manager, F5).

> When I work on my computer, I do paragraphs, abbreviations, full stops and you have spell check which helps because not everyone is 100% on their spelling (Retail assistant, F3).

The manufacturing employer (Emp 8) also appreciated technology: "You know, I think the level of your English, your written English doesn't determine how successful you can be in an employment setting. Technology is a wonderful thing, you know there are so many tools out there to assist people, to help people."

Nonetheless, computer editing did not help improve language according to the university lecturer (M1). He reported that: "You know when we get some senior graduating students to write reports in a way where they can edit and edit to get the spelling correct and they ask Bill Gates for grammar

suggestions, it's never perfect and it's often quite shocking" (University lecturer, M1). The Master's student who was a manufacturing manager had the same viewpoint. Describing typical workplace language she encountered on a daily basis, she said: "Even the emails are either strictly bullet points, they're not coherent and there's no cohesion and the spelling's atrocious and we've got spell check." (F6).

Interestingly, the manufacturing employee from the Philippines (M FNNS) noted that technology may have led to a decline in the language skills of native speakers: "Because gadgets are everywhere and they (native speakers) do writings in computer, in a cell phone, not in a hand writing. So that's what I believe. That's why they get the poorest rating in writing – because of the technology and also the slang".

5.2.4 Summary

What forms of English are generally used in the contemporary workplace? According to my research participants workplace tasks generally revolve around using templates, forms, policies and procedures and health and safety notices. Technical and work specific jargon are used in most industries, but this jargon (including codes and acronyms) is used across languages.

5.3 Constructs that are valued in the workplace

As regards the constructs that are valued in the workplace, there are two questions which are appropriate for test validity purposes: Are the constructs being tested, those constructs which are used and valued in the workplace? Is the test measuring construct-irrelevant language abilities? (Knoch & Macqueen, 2016)

5.3.1 Communication in the workplace is valued

Language test developers may assume that employers want employees who can express themselves precisely and proficiently in both an oral and written format. What transpired from the analysis of the interviews, however, was that communication took precedence over linguistic proficiency. This may reflect the fact that skills identified as being important for workplace used to be the three Rs (reading, writing and (a)rithmetic) but these have now been replaced by the four Cs (critical thinking, creativity, collaboration and communication). A comparison can also be drawn between taking a language test, which assesses an individual's linguistic output, with language use in the workplace, which is collaborative.

In education, accurate reading and writing skills were not regarded as being as essential as communicative competence. The university employer (Emp 1) put it this way: "despite everything, a

teacher's main resource remains their voice and your ability to communicate with students" while the primary school principal (Emp 3) outlined the skills of a good teacher thus:

> You know, to communicate effectively with the children and of course with the parents, with colleagues is really important, and then it's the written skills of being able to communicate clearly your thoughts via reports or it might be regular newsletters, classroom newsletters or emails out to parents. Those sorts of things are really important ... My advice for them is to remember what standard is required for primary teachers and that primary teachers are not writing academic essays (Primary Principal, Emp 3).

Similarly, at tertiary level communication seemed to be paramount. The university lecturer test taker (M1) described a good economics student (or lecturer) as being able to communicate effectively:

> There's no question that good communicators are in some sense better economists. And the same thing holds for students. A student who can communicate their ideas well will do better because if you can't communicate your ideas, it's difficult to know on paper what ideas you have (M1).

After taking the IELTS test, this lecturer reflected that the test did not appear to be testing what he believed it was testing, that is communication:

> Again, the test isn't testing the functional ability to communicate that is required for the job. So, there's a standard and then there is what actually works. I'm assuming these people aren't actually killing people because they can't read the instructions. Yes, it's a test that is not matched well to the purpose of the actual thing that you are trying to test which is effective communication for the purposes of whatever the context is (M1).

For the IT and business employers, intelligible communication was regarded as being paramount for the workplace, together with the proclivity for employees to question what was happening around them to prevent problems being exacerbated. In fact, many participants compared the communicative aspect of workplace practices with essay writing (an individual practice). The IT employer (Emp 5) noted that "We must have a pretty low bar because, in all honesty, they don't need to be able to write essays. They just need to be able to follow instructions and communicate ideas" while the business employer (Emp 7) contrasted communication with proficiency: "But what I think is important for employers is how can a person communicate in the workplace in a way that everyone is going to understand ...Today it's about communication, isn't it rather than proficiency" (Emp 7).

Apart from communication skills, problem solving skills were more important than accuracy in language for all the employers I interviewed. This was epitomised by the manufacturing employer (Emp 8) who valued day to day problem solving skills:

> So when things are in control it's easy but when things get out of control that's when we see the risks for health and safety issues, the risk for quality issues so for me a competent employee is not only understanding what they do day to day but it's a bigger picture, it's being able to use some common sense and understand, hey, something's not right here, stop, what will we do? That to me is the main comprehension.

Likewise, the ability to collaborate, question and note down problems was commented on by the health and safety consultant (Emp 10) who described a competent employee in the following way:

> A competent person can just, you know I say to them you're not writing down an essay, you just write down the date, the machine and something, and a couple of words and I said that will then trigger me to do it all. So, they are competent to say that, to write it down and they're not worried that I'm going to say 'oh my golly, look what you've done, that's a double negative' you know. So as long as they can write down pastry brake failed or mixer broke. Even if they can write down a limited amount, that's enough to trigger and we can do all the investigation later. But competent people can write that down. It's when something goes wrong.

All participants (test-takers and test users) claimed that they may write short reports but not essays. The ESOL teacher (M4) explained that this is especially the case for people working in technical areas: "My son who is an electrician doesn't need to write essays. He needs to communicate." For the manufacturing employee from the Philippines (M FNNS) the main skill at work was communication with colleagues and bosses. He believed that he communicates well at work, in terms of listening and speaking. As for reading and writing:

> We don't usually do much reading at work because our job is to install panels and to communicate with our bosses and our colleagues and reading drawings but that is totally different to reading stories. The drawings are what we are doing, and all the details are in there ... And writing, we don't usually do writing at work because if we want to say something to our bosses or to the office, we just send them an email, just two to three sentences according to my needs (M FNNS).

This particular migrant noted that there are four Filipino workers in his building fabrication factory and that although he gained the highest grades in IELTS, the other Filipino workers "can still manage to communicate with our bosses and do their job properly so I don't think there are much links between the scores and the job" (M FNNS).

Pragmatic competence in English was deemed most important by the Chinese businessman (M CNNS) for living and working in New Zealand. He observed that he communicated well with his colleagues, customers and business partners because he had attended life skills classes in New Zealand whereas, for him, candidates who may score well in IELTS:

> They don't know more about New Zealand culture and like the Kiwi speaking, but their English skills are better than me because they study them at school. They have the skills for sitting the exam … I have no problem (with pragmatic language competence in NZ) because I went to a New Zealand life class ... I think every immigrant should take a New Zealand lifestyle class, you know. That's more important because that's why they are staying in New Zealand. They've got to learn how to survive (Chinese businessman, M CNNS).

5.3.2 Collaboration

The benefits of a collaborative team environment for communication purposes was referred to by some participants as being important. The IT employer (Emp 5) appreciated employees' experience using technical language and a team environment: "You need to be able to communicate technical concepts and you hope that most of them have worked in distributed technical teams so that they've had experience where they've had to apply that language". She argued that the collaborative nature of IT work meant that listening and speaking skills were paramount:

> It's nuts, I don't think you can set a higher bar for some and you know this is actually the most important part of the test (speaking and listening) in terms of working collaboratively and actually being able to fit in to the society and being able to assimilate and make bonds and form connections. It's the speaking and listening that makes it. When do you ever write this sort of stuff? (IT employer, Emp 5).

The telecom technician (M2) mentioned that he had always worked in teams with a wide range of nationalities and that this collaborative environment helped overseas employees communicate with New Zealand clients:

> But in a lot of cases with people who speak a second language - presumably they're not on their own. They're in a team of people who can communicate with the locals, so they would be able to do their job cos they can communicate with others (M2).

The retail assistant (F3) recognised that many overseas workers were needed for the Canterbury earthquake rebuild but, as she noted, they do not work in isolation: "Builders would look at architectural designs and stuff like health and safety … I guess because they work with architects, so they work on instructions … But, it's just a big collaboration".

5.3.3 Sophisticated versus simple language

The non-native speaker test takers did not discuss the importance of sophisticated or simple language in their workplaces; rather, they commented on the lack of sophisticated language or accurate language produced by their fellow native speaker work colleagues. However, the perspectives of employers regarding language change in the workplace proved to be quite pertinent for the argument about whether test responses and hence test score results reflect language constructs used in the target domain. Listening and speaking were consistently identified by test takers and test users as being most relevant for communicative competence in the workplace. In this section it will become apparent that the participants' comments on the use of simple or complex language in the workplace may have some bearing on the relevance of gaining high test scores in the reading section. However, the points participants raised have a stronger significance for the testing of writing. As a teacher of English and a language testing examiner, my purpose has always been to 'spot' inaccuracies and help students develop proficiency. Language teachers in general want to enable students to improve the breadth and depth of linguistic skills they can employ. Yet, the native speaker participants in my research (test takers and test users) argued that language use in the workplace has changed; clarity and simplicity are now valued over linguistic sophistication.

As regards reading complexity, some of the participants remarked that texts are meant to be comprehended by teenagers. The primary school teacher (F1) felt that the IELTS texts she had read were far too complicated: "If you compare those readings with a newspaper article, which is what I think they should be using. Newspapers in New Zealand are presumably written for a reading age of 14." The accounts manager (F5) was not sure of the average reading age in New Zealand but knew that they had need to adapt their practices at work to meet this age:

> So, if we're saying that the average New Zealander has a literacy level of 15 or 16, so when we're doing the e-learning stuff, they say you have to focus on that level. So, you have to

> make sure a 15-year-old can understand it and whatever the literacy level is in New Zealand and it's not that much higher, I don't think. This is pitched way above (F5).

As far as writing is concerned, the participants' highlighted a more dramatic change in language use. In the business world, training nowadays encourages the use of simple language as described by the business consultant (Emp 6) who regularly runs courses in this area:

> Plain English is absolutely the mode for employers these days, even in business reports etcetera. You know, we used to write like academics in the workplace but as a professional, plain English is absolutely the mantra. We do a lot of training here … and all of the business writing courses drum that into you – if you say it simply and still get your message across, do it that way.

This could be because people are stretched for time in the modern workplace, as assumed by the immigration business director (Emp 7):

> Well, you see the reality of the workplace now is that everything is fast and less is more. So, you know, in the old days when we used to write submissions to the government for example, it would be pages and pages and pages of detail. It's all gone now and we're writing in bullet points because you're busy on the other side of the desk and you don't have time to read a novel. So, you know, I know that's probably the requirements of the test but maybe that's not actually the real world because the real world is in less can you get the message across.

Although many employers liked well written English, they conceded that this did not translate into competence: "I'm a bit of a grammar Nazi but I understand that it's not important for everybody in their workplace. Unless, that's part of your work, you don't need to be perfect in writing. Some of the best people in the workplace are not." (Emp 6). Similarly, although the IT employer (Emp 5) prided herself on having 'good' English, she commented that in the IT field casual communication was standard: "There's not an expectation that anybody is using formal business language and most of the communication is email so it's quite casual."

These opinions were reflected in the comments made by the native speaker test takers who worked in the business/IT and manufacturing fields. The accounts manager (F5) worked on e-learning packages for businesses across New Zealand. She reflected that in her company:

> Like, even in my own world, I write a lot of RFQs for government work and even like 3 years ago when I was writing RFQs, a sentence would be four paragraphs long almost and they

> would have all these complicated words but even now it's all about simple language, like the way we do it has changed dramatically. It's not about long words that would take you about half an hour to work out what it means, it's all about really simple language. So that's a response for a big proposal for a government agency, so we might be building e-learning for any government department. Once upon a time it was all about really complex language but now, they want simple language. There's been a real movement, I think.

This change in language use was also described by the Master's level student (F6) who worked in manufacturing. As she pointed out, overly sophisticated language could lead to comprehension problems for native speakers as well as non-native speaker staff members:

> It's pretty much keep it simple for the manufacturing world. If you keep it simple, keep it straightforward, keep it easy-going, keep it easy to understand, then mistakes are not going to happen. If you use complex, fancy words, big long words, big long explanations, basically then the meaning is not always going to go in and people are not always going to understand, even native speakers.

Comparing the test requirements with real-world, authentic language in use, both the business consultant (Emp 5) and the ESOL teacher (M4) believed that describing a graph would be used for business presentations where clear language is paramount. According to the ESOL teacher:

> Part One writing for me is for the business world, so the sentences need to be short sweet and to the point and those are the ones that need high marks. That one there should get an excellent mark (it was rated as a 6.0) but the other ones in terms of 7.0 or 7.5 are getting too complicated ... You see, I would argue that this type of writing here is the sort of stuff that I would write for speech notes for politicians because they only want complicated sentences that no-one can understand (M4).

This situation appears to be no different in the education sector. Participants noted changes in society's expectations at primary and tertiary level and wanted to know if language tests were keeping up with the ways in which the use of English has evolved in schools and tertiary institutions. The primary school principal (Emp 3) remarked:

> But society has shifted … I talk about commas but commas aren't used as much nowadays as they used to be and so I guess, language changes. Has the marking of that test, is it still relevant to 2018 New Zealand society?

From a student's point of view, lecturers did not expect complicated language either. The university student (M5) explained that:

> I know that one of the requirements in writing a physics report is to use simple sentences. Like, they don't want you to use confusing language at all, and that's university physics, so that's almost the opposite to what they're testing for here ... Maybe in other majors they may want other stuff, but I still feel that in general they want you to show understanding of something and using big complicated words, using really hard English doesn't show you anything about a subject. It just shows you can do English for the sake of doing English.

This participant was an undergraduate studying computer science, which may not be regarded as a language rich subject. Nevertheless, even when writing a PhD, the university employer (Emp 1) whose doctoral thesis was in the area of education, argued that simple clear vocabulary was best:

> Why use huge long words when short specific things convey the idea and the information so that it is accessible to the reader ... I don't mind having my vocabulary expanded, but if I'm having to go into a dictionary to find out what you actually mean by this, it's not communication. What the clarity is, that's surely got to be fit for purpose.

The business consultant and business director summed up the conundrum between language for tests and language for the workplace in the following way: "I think it's measuring proficiency according to, and it's a bit subjective, but according to something that was never made to be a workplace measure." (Emp 6)

5.3.4 Accuracy

Language expertise is measured in terms of breadth and depth of lexical and morphosyntactic output. Accurate English, produced spontaneously without time for proof-reading, is essential to gain high scores in language tests. Yet, the business employer (Emp 7) raised the point that while language in society used to be mostly accurate, today this is often not the case. Talking about the local newspaper, he believed:

> Before smartphones we would get *The Herald* and everybody would read it and back in the old days, it was always accurate….and now when you read it on your smartphone, you know, when you read things, there's so many mistakes in there. There are sentences that sometimes don't even make sense. You have to re-read them to know what words they've put in wrong because it's quick.

Statements made by all of the educators may surprise the creators of IELTS tests and the marking rubrics because the education employers were not overly concerned by what they labelled as 'mechanical errors' such as spelling and grammar. The senior educator I interviewed had been a marker for NCEA level 3 (university entrance tests for secondary students in New Zealand) and claimed that when assessing a piece of writing "We read that in terms of the sense making that people have. We don't read it for accuracy of language whereas this (IELTS) is straight accuracy" (Emp 2).

The primary school principal (Emp 3) explained that any work is double checked collaboratively before it is released to the general public because teachers are not invariably accurate in their use of English.

> It varies, so for all of our correspondence that goes public … it's always proofread and double checked, including mine so I would never send anything out that's not checked, simply because there are bound to be errors and yet the standards for teachers and principals is high as a society. You know there's nothing worse than a school newsletter that goes out and it's got spelling mistakes in it. And you know we don't proofread our own. Once you've written something, you know, you can't proofread that properly … it's always easier to proofread someone else's work.

Likewise, the university lecturer (M1) worked in an international department at his university. He reflected on the purpose of written communication and how the role of the reader affected the importance of accurate language:

> We have 23 people working for me and probably 17 from different countries, so there are 17 foreigners at least and they don't all speak perfectly. (*English not their L1*). The emails I get, a lot of them are like that but if this was going to go to the Vice Chancellor, we would tidy it up but if it was an email to communicate an idea to me, this would be perfectly fine.

Outside of the educational sector, in IT and business employers valued employees' ability to collaborate over grammatical accuracy. The business employer (Emp 7) employed many nationalities as immigration advisors since they needed to interact with migrants who spoke a language other than English. Yet, as this immigration director (Emp 7) pointed out, non-native speaker employees were able to double check the accuracy of their English with native speakers of English if the correspondence was intended to be read by government officials "The benefit we have, obviously, is that we have native English speakers so that any of our written work, before it hits the government, can be double checked."

Similarly, the IT employer (Emp 5) explained that she valued communication and collaboration more than grammatical accuracy:

> I don't think there is an expectation that English is going to be perfect in the written communication. In all honesty, it's more if they are able to understand and follow instructions, ask questions and communicate information. It comes down to if the information has been presented in a professional way ... that to me is just as important as whether the grammar is absolutely perfect.

Furthermore, for many participants, accurate English was irrelevant and did not equate to competence, as exemplified by the Master's level student (F6) who was a production manager:

> Grammatical errors, and all the rest of it. In my profession, I wouldn't even look at or pick up on. I'm reading for meaning in the workplace. I want to find out what I need to do, what's this person trying to tell me, what's happened and how can I fix it. The spelling and grammatical errors are irrelevant and that's across the board.

The business consultant (Emp 6) echoed this view, saying that grammar is not "important for everybody in their workplace. Unless, that's part of your work, you don't need to be perfect in writing. Some of the best people in the workplace are not"

5.3.5 Summary

The constructs which are necessary to display workplace competence include the ability to communicate and collaborate using clear, simple language. Complete accuracy and sophisticated language are not expected.

5.4 The literacy skills of native speakers of English in the workplace

The perceptions of my research participants regarding the linguistic skills of native speakers of English in the New Zealand workplace contribute to the extrapolation inference since they help paint a picture of the nature of general language proficiency in society. Further, the perceptions of native speaker employers highlight how decisions are being made regarding test score results in New Zealand Standards of test performance required by policy may only be ethical and justifiable if test scores relate to native speakers as well as non-native speakers of English. Secondly, decisions are beneficial for non-native speakers, as well as native speakers of English, if the test scores relate to language use in society.

5.4.1 The perspectives of native speaker of English test users (employers)

Starting with the education sector, the employers' estimations of the English language abilities of native speakers they knew or worked with, varied widely. Some predicted that native speakers would gain a score of between 4.0 and 5.0 on an IELTS scale whereas others assumed that native speakers of English would naturally perform well (8.0 – 9.0). Those who estimated that native speakers would gain the highest scores possible included the primary school principal (Emp 3) and the senior educator (Emp 2). However, the language that they used to describe the attributes of people who would gain these scores revealed the need for language assessment literacy. The primary school principal (Emp 3) initially believed native speakers would be likely to score 7.0 if they had issues with language but 9.0 if they had reasonable language skills:

> I would think that New Zealanders would get, well New Zealand born if we use that term, they would get 7s, possibly 8s. You know I have to talk to some of our so-called New Zealand teachers about pronunciation and words and modelling those. You know 'somethink' because that has an impact on spelling, so they may say that doesn't matter but it does … I'm guessing that with New Zealanders, I wouldn't be surprised if there were some 7s, but I wouldn't expect all 9s for example.

The senior educator (Emp 2) estimated that New Zealand teachers would score slightly lower "Lucky to be 7.0 sometimes, 7.5". Interestingly, the Education Council of New Zealand expects teachers who are applying for registration in New Zealand to have a minimum of 7.0 in all skills. This educator (Emp 2) supported her viewpoint by saying "I think some of them don't understand language and how language works in the way that when English is not your first language. Often because you've had to study the language you understand the language in much more formal and complex ways than native speakers do."

Nonetheless, after seeing a sample test and sample test responses, both the principal and senior educator revised their hypotheses, as exemplified by the senior educator (Emp 2):

> I suppose it re-established my own thought that probably native speakers wouldn't do that well in it because knowing how they use language and having been involved in literacy teaching as well, I know some of the strategies of even teachers are really, really poor ... Straight up, I know some of the teachers I've worked with, you showed me those reading tasks, they would not …they would be those 5s and 6s. Well given that native speakers are coming in as graduates with literacy at level 2 NCEA. It's actually a far cry from 7.0 that's there.

The university employer (Emp 1) had had many dealings with students both at secondary and tertiary level. His conjecture on the ability of native speaker students and teachers was more pessimistic:

> Now, of course we know that with native speakers that can be hugely variable. Having been a Dean at a school and having to correct some of my colleagues' English. I see that with our graduates, even now ... I would suggest that many native speakers and particularly if you have high school leavers who have not had, like they have an option to bail out of English when they start having elective courses so if you have people who really have felt that they have struggled in English as a subject at school and have acquired the basic literacy credits for UE (University Entrance) and maybe in engineering or maths or science or whatever, many of them would be low, I'd say in the 4 – 5 band.

As he pointed out, people who work with numbers and symbols are not necessarily proficient users of written English. The ESOL employer (Emp 4), who came originally from the UK, argued that many average New Zealanders and British people in the workforce would score as low as 4.0:

> I don't think they would do very well because I think it's very complex language based, quite technical and I think most Kiwis don't, well you know you forget the grammar you learned at school and I don't think they learn much grammar at school in a formal way nowadays so I don't think most Kiwis would get a high score ... I think most Kiwis would write worse than that … let's face it Kiwis spelling can be appalling. Well probably British as well but I mean I haven't been in England for a long time, in terms of working.

When I queried if she meant blue-collar workers, she seemed to believe that this applied to most average people in the workforce:

> No, I mean anyone. Oh, I imagine that depending on education level, you might get a very low. I don't know how low a 4 is actually, but we do have people who have quite good English and who have only got 4s. They're learners but their spoken English is good, so you can't necessarily tell how good the other parts are (ESOL employer, Emp 4).

In the business and IT sectors, employers were of the opinion that literacy rates for New Zealanders were not invariably high. For this reason, they argued that native speakers of English may not fare as well as non-native speakers in a test of this nature. The most striking comment was made by the business consultant (Emp 6):

> In certain industries [native speakers of English would gain scores which are] lower than our migrant workers. We know through research that there's very low literacy rates across the New Zealand workforce, and in fact, across the New Zealand population ... and I've seen it myself. When I go to train, where often young people, not always young, but often young people. When I give them the feedback form, which is very simple – it's just what did you find useful about this workshop – they cannot read some of that. Now, that's New Zealanders so that blows me away.

The business employer (Emp 7) lamented that nowadays native speakers are not aware of how to produce accurate English. He argued that New Zealanders would gain the same scores as most migrants due to mistakes in their written output:

> And the risk is that as an employer, I'd make an assumption that because you're a Kiwi, you're going to be able to write as well as anyone else here, but in fact you may not be able to … You know, I'll get somebody's CV in and I'll look at that and I will say I found ten mistakes so I would be worried if that person could actually be that accurate in the workplace, but that person could be a Kiwi, but he still makes mistakes.

According to the IT employer (Emp 5), the way many people write in the workplace has changed. She compared the written responses with how members of her family wrote. Describing her children, she echoed the thoughts of many of my participants that young people are not taught how to write effectively in today's schools, although as she pointed out, this has no correlation with their cognitive abilities:

> They know how to write but not particularly well. You know they're getting better with it. They're both 16 and they're both very bright. I've got a boy who will probably go on to be a computer science student and he will probably get the exact same scores as this guy (M5). You know he got a merit in English last year, but I don't know if he'd be getting a 7 in this ... And then the girl who works for me here, she has no idea how to use apostrophes, so she wouldn't do well either. It's ridiculous (Emp 5).

Many participants reacted strongly to reading the response written by the accounts manager (F5, who scored 5.5), explaining that this piece of writing was most similar to 'natural' English in the workplace: "It's really funny because it's just the way my kids write. The whole thing with writing just the way they speak, and this is exactly how they would write. Like they're writing just as if they're talking to you." (IT Employer, Emp 5).

When interviewing the employers in the manufacturing and retail sectors, it became apparent that they did not expect New Zealanders to do well in this type of test of English because: "this test is too hard, too harsh yeah, and some people would actually struggle with that or most in our industry and that is Kiwis as well as internationals" (Employer 9, retail director). Talking about her husband, the IT employer (Emp 5), compared his language skills with others who work as tradespeople:

> And you know, this level of written English, there would be a large sector of the population that would be writing English at around this level (IELTS 5-6). You know, if you asked my partner to sit down and take this test, he would be using very simple basic English because he's a tradie.

The health and safety consultant made pertinent remarks about people who were generally employed in the food manufacturing industry. She pointed out that the non-native speakers were often tertiary educated and had better language skills than native speakers of English. Yet, staff in this sector were not employed for their English language skills, and complex language was not expected:

> And the Kiwis, none of them would have passed that, absolutely none of them. Maybe one woman, we've got one woman and maybe she would have passed but she won't stay here that long, I know, she'd be as bored as anything. But the ones there, when I do training, I have to so bring it down. You know I can't use any complex language. And most of them had a hard time at school so the general literacy of the Kiwi workers is low (Emp 10, health and safety consultant).

Another issue raised by this employer was that that the majority of people she worked with were adept at their work but "as soon as you put a piece of paper and say, 'Can you read this?', they shut down." (Emp 10). This may be due to the fact that some workers have learning disabilities and therefore struggle with literacy:

> I find it really hard because quite often in a food industry you pick up a lot of people who have learning disabilities. You know, I used to just assume that they could read the sentence until I did a lot of training and realised the reason why I was the quality manager was because I could read and understand and then when I gave them tests I was, you know even though they were Kiwis, I would still read the questions out and then we got more of a success rate and then people would say well that's cheating and I would say, well there's two skills here, there's reading and writing and you know some people can't (Emp 10).

Yet, native or non-native speakers who struggle with English may be very competent in the workplace. As the health and safety consultant (Emp 10) perceptively explained, the workers in her sector may not have the best English; yet, "they know their machines, they know the sound of it … they've found their niche in society, and all work is noble".

5.4.2 Perceptions of native speaker test-takers vis-à-vis native speaker proficiency

After taking the test, the native speaker test takers suggested that the majority of New Zealanders would receive similar or lower scores than they had. The older participants (M1, F1, F2 and M4) argued that younger people would find this type of test particularly difficult. The ESOL teacher was of the opinion that teenagers would score as low as 3.5 to 4.5, although the test score results, especially for the reading section, do not support such a response. The phrases 'have trouble with' and 'struggle with' were often used to describe how native speaker test takers believed New Zealanders would find this type of test.

The primary school teacher was of the opinion that school teachers would do well as long as they were informed that they had to follow the instructions exactly: "Yet, "10 – 20 percent would have trouble (with the test). And the point is, does that necessarily make them a bad teacher?" (Primary school teacher, F1). This echoes the perspective of the health and safety consultant (Emp 10), mentioned above, who saw no correlation between English proficiency and workplace competence.

Talking about students at tertiary level, the lecturer (M1) felt that most native speaker undergraduates would not do well:

> If you took the English-speaking students that I have in my first-year economics, they would struggle. There is no question that they would struggle. If I sat them down and told them that I was going to mark this on some NCEA scale of writing ability, they may write differently than what I get on the questions that I ask them, so I can't say for sure, but many of them would fail horribly.

As regards the IT sector the telecom technician thought "Telecom workers would really struggle" (M2) while the ICT manager was of the same opinion "People in my team in my current job would have struggled with this" (F2).

Discussing the New Zealanders that she works with F6 (Master's level student) observed that even upper management would not gain good scores:

> Basically, they don't have the level of education needed to, or the level of language needed linguistically to do well in this test. I struggled and I'm doing a Master's. You get the

> commonplace Kiwi who's finished high school at 15 to go straight to work, there's no way they'd have the level of language to do this test. People I talk to on a regular basis, no way they'd be able to pass. My CEO in my company would probably only get a 4.0 in the writing. I've seen his emails and spelling mistakes galore plus lack of cohesion – sometimes it just doesn't make sense. It's pretty much bullet point, bullet point, bullet point ... No full sentences. I don't think they'd be able to write a paragraph and that's at upper management level. I don't think they could write a coherent sentence if they tried. And I deal with management all the time (F6).

For the native speaker test takers, the majority of New Zealanders would, therefore, need training to attempt this test even though the test purports to be testing their English. As the primary school teacher declared, "Oh I think most New Zealanders wouldn't pass that, not without coaching beforehand, being taught to the test" (F1). The university lecturer discussed how the average New Zealander would fare:

> I suspect that if you asked someone to write 150 words, man on the street sort of thing, that they would really struggle. Interpreting graphs ... if they don't understand pie graphs maybe they wouldn't be able to write anything but that is not because they cannot write but it is because they don't understand that particular task. But, I think that people would struggle to write 150 or 250 words if it is not something that they have practised and prepared to do (University lecturer, M1).

5.4.3 Perceptions of non-native speaker test-takers regarding native speaker proficiency

Although a few non-native speaker test-takers felt that native speakers should do well because the test is assessing their L1, it was surprising for me that the majority of the non-native speakers did not believe that native speakers of English would necessarily gain high scores in the reading or writing sections:

> I think they would score well in speaking and listening. I think they might because they use it every day – not so much in reading and writing. Well, just what I've seen and what I've observed from being in the workplace. So, the first few months I was here, I worked in the aged care industry and then I moved to the labs. You know, I've had nurses who were native speakers who can't even write well in their nursing notes … And some of them have good sentence structure. It depends on the person, but some of them don't and even with

> spelling. For us, spelling is supposed to be a good marker of English in the Philippines but here some of the scientists can't even spell (F FNNS).

Likewise, the non-native university lecturer (M GNNS) mentioned that spelling, as well as syntax, was regarded as important when he was learning English as an L2 in school. He was bemused that this was not the case for his New Zealand students:

> It doesn't seem to be a given for people coming out of high schools in New Zealand to have high written literacy skills. When I consider my English learning in Germany, we were marked down for every spelling mistake and every grammar error and so we could only pass if we had, on a page of writing, less than five grammar or spelling mistakes, which many of the work that I have seen from native speakers does not fall into this category.

This lecturer from Germany expressed the same view as the native speaker lecturer (M1) because the native-speaker of English undergraduate students that he encountered had a wide range of English literary skills, with some being proficient whereas others had what he described as "challenging" levels of English. He (M GNNS) believed many New Zealand students would score 6.5 or below:

> Especially here in New Zealand, I come across quite challenging standard of knowledge of English grammar and writing. Construction of arguments within tertiary students, it's often challenging. Considering that students within a university environment have problems constructing proper lines of argument and writing grammatically correctly. Punctuation issues all over the place, I don't know how much we can expect from foreigners ... Because of the quality of the writing, despite being native speakers, especially in teaching degrees, it's something I cannot pass. Just because the clarity is not there through poor writing (M GNNS).

This viewpoint was echoed by the university student from Singapore (M SNNS) who commented that New Zealand students did not appear to know how to structure their academic work: "I'm not trying to look down on Kiwi education but when I talk to my classmates, it seems that high school and primary school does not prepare them for full-on academic writing in university."

Similarly, the primary school teacher from Germany (F GNNS) who had done all of her teacher training in New Zealand, suggested that some of her native speaker colleagues at university, and in the schools which she taught in, struggled with English. Talking about her partner, this non-native

primary school teacher argued that he would most likely score lower than her because he is a dairy farmer with a word budget:

> He has a word budget and his budget is 100 words a day and that's it so maybe a 6.5 or 7.0 (for speaking). He'd maybe get an 8.0 – he'd be ok in listening and maybe 6.0 in writing. And, in reading, he wouldn't score very high either. He doesn't read really – he looks at things so maybe a 6.0 or a 7.0 (F GNNS).

For the non-native speakers who worked in business and manufacturing the estimates of the scores that native speakers they knew would gain was even lower, ranging from 4.0 to 6.0. They brought the debate back to academic proficiency, pointing out that in many cases only well-educated native speaker test candidates would gain 7.0 or above.

> I think it really depends on who they are. If they get well educated, maybe they will get over 6.0 in the four skills … but for a person who cannot get well educated, I think they can just get 5.5, because you know it's the same in Vietnam. If you get a person who don't study well, you know Vietnamese is their native language, so they can just talk but what they talk is impolite. And IELTS test, just talk about language that is polite and formal you know (F VNNS).

Similarly, the Chinese businessman had given a test to his homestay mother when he was studying English. He declared that she could not understand the test and would, therefore, probably gain a score of around 5.0:

> My homestay would get 5.0 "She said oh no, it doesn't make sense" so I believe she would only get 5.0. I believe 6 or 7 is a fair mark for native speakers but anything above 7.0 is more academic. It's more education levels. I have a situation where Kiwis ask me how to spell a word. That's normal. So, in that case it could be listening, speaking, reading is no problem, but writing may be. It's like if a Kiwi writes a paragraph for an insurance claim and you show this to the IELTS examiner to see how many points they could get – it could be 4, 5, 6, 7 but above 7, this guy would be really academic (M CNNS).

Comparing university educated workers with school leavers, the production manager from India suggested that school leavers would find it difficult to complete this type of language assessment, especially in the writing section:

> It's hard to say but in my opinion, native speakers don't concentrate much on the grammars and all that. If they go for like academic (test), maybe 5.0 or 5.5. It depends on the person,

> like age, what they are doing. But if they are uni student and still studying they can easily get 6.0. If someone is doing PhD and native speaker and like his English is really good, he can get 7.0. It depends on the person, what native speaker ... This is why I'm telling you, in my workplace, mostly my co-workers are not very well educated. It's like a labour work for them. I think they just leave their school, so I think that in that case, I don't think they can do these kinds of task one and task two (M INNS).

The manufacturing employee from the Philippines reiterated this viewpoint because he had not seen examples of accurate writing from his colleagues:

> In my observation, my Kiwi colleagues are poor in grammar and writing. They are not good in spelling. I'm not saying in general but the people I have been working with, I think they would get about 6.0 for their writing because I have observed them with writing and spelling and there are times when I believe I am better than them. Sometimes even the British (M FNNS).

However, as the librarian from Poland (F PNNS) pointed out, it was often acceptable for native speakers to make mistakes with their English but not for non-native speakers: "Well, I would say that sometimes when I read some emails that Kiwi people write, I'm thinking, this is not right what I've learned like 'do not' or the commas are not right so I just think well, they're allowed to make mistakes and they do"

5.4.4 Summary

Although some native speaker employers initially expected English speakers to do well in an English language test, when they viewed a sample test and test responses, they changed their minds. In general, the majority of my research participants did not expect New Zealanders to fare well in this type of test in the reading and writing sections. The ability to gain high scores was equated with higher language cognition gained through extensive use of English in academic or professional circles. However, most speakers of English have basic communicative competence which is generally demonstrated through their speaking and listening competence.

CHAPTER 6

FINDINGS PART THREE: PERCEPTIONS OF ENGLISH LANGUAGE ASSESSMENT AND THE IMPLICATIONS AT A SOCIETAL LEVEL.

6.0 Introduction

Participants in this study commented on the social, economic, societal and political consequences for New Zealand. Test takers' and test users' perceptions of the wider societal impacts of using the test as a gatekeeping device (Macqueen, Pill & Knoch, 2015) for migration and professional registration in New Zealand helped build a validation argument for RQ4: What do test takers and test users perceive to be the consequences of test use for the New Zealand context? According to Kane (2016) the evaluation of consequences, both intended and unintended, plays a crucial role in validating score uses and show whether a high-stakes language test is being used ethically in the context for which it was intended to be used.

6.1 Social consequences

6.1.1 Importance of the New Zealand context (the target language use domain)

The majority of participants in my research queried the value of using test score results from an international language test in the New Zealand context. As Kunnan (2018) suggested, "assessments ought to be evaluated in terms of whether they can bring about beneficial consequences to test takers and the communities in which they are used (p.200)."

The business consultant (Emp 6) highlighted that employers need employees who can assimilate readily into the New Zealand workforce but that this type of test was not beneficial in this respect:

> Most of my training is focused on training New Zealanders to understand how different we are when we communicate. So, it's not just how quickly we speak and our accent, which are two huge barriers to newcomers, but it's also how indirect we are. It's using something like 'we' instead of 'you' when we give instructions, how we're pretty fluffy with our feedback sometimes. So, there's so many nuances and IELTS doesn't capture any of that New Zealand English stuff. Our slang, you know, that's huge. So, for employers, they want to know that their employees are work ready in a New Zealand context and it's pretty rare for someone to come out of IELTS and be able to say, "I really understand now how to work in a New Zealand environment".

6.1.2 Social consequences for the New Zealand context

After taking the test, there was general bemusement from the native speaker test takers and test users regarding the test, why it was being used in New Zealand, and the range of occupations that it was being used for. One consequence of using this type of test may be that it can exclude people who might be able to usefully contribute to society in spite of issues with literacy or language proficiency. Indeed, many participants argued that society needs a wide variety of people, and that since the test appeared to them to be testing educational levels, over-using it could lead to an unhealthy mix of people with inappropriate skills sets. The accounts manager (F5) perceptively pointed out:

> I'm thinking we're making it so hard for people, ridiculously hard. Like, we don't want all professors in NZ. We want builders and we want florists so by making it so hard, we're eliminating a huge proportion of the population, surely ... I'd like to see more flexibility, depending on the occupation and the person. I think there's a lot to be said, apart from how you read and how you speak, to bring you into a country. Like you could bring so many more skills, like you could be the best painter in the world or the best pottery … you could be an artist or music for example that would bring so much to New Zealand. Just because you can't write an essay should not exclude you. I think there's a much bigger picture than how you can speak English.

The IT employer (Emp 5) argued that this type of test appeared to be assessing high levels of language proficiency, but that in her company, each employee had different skills, and that all of these were valuable:

> But you know it's what are we looking for in this country, what's going to be valuable for this country, do we need novelists and journalists or people writing research papers, or do we need people who can, like in this building, write code? I've got great English skills, but I can't write code. It's really interesting – I wonder what other countries do?

The concern for the manufacturing and retail employers was that a focus on language assessment would lead to a lack of workers in their sectors "and if that's what we're going to get, we're not going to get the factory workers." (Health and safety consultant, Emp 10).

The primary school teacher (F1) went further, relating the need to follow instructions in the test to the possibility of attracting migrants who were socially inept:

> It seems as if they are trying to find a very narrow band of people and that very narrow band of people may not be exactly what the country needs. They are going to find academically inclined - you could well end up with an entire cohort of high functioning autistic people who follow instructions precisely and are well able to deal with all the academic things but can't interact socially.

This same participant (F1) suggested that it was important to foster a society where different learning styles were valued and that a language test may not, therefore, be appropriate:

> How's someone with dyslexia ever going to get entrance to NZ? So that means when you think about it that Richard Taylor who runs Weta workshops is not going to pass this test because he has dyslexia. All these immensely creative people wouldn't pass.

The librarian from Poland (F PNNS) thought that New Zealanders with dyslexia may be intelligent but would struggle with an IELTS type test. Moreover, non-native speakers with dyslexia would find it even more difficult:

> Well, if you are a mathematics or physics teacher, what does it say because there are many Kiwi who I know, and they have a Master degree and they are dyslexic. So dyslexic people and they are genius about digits. So that is irrelevant, and they would just not go there if they are non-native speakers because they would fail terribly, probably a 3, if they are dyslexic. If Kiwi who are genius in mathematics, in physics, they cannot pass this one, so how about non-native speakers?

The consequences of using a language test for people with dyslexia was also discussed by the IT employer (Emp 5) because she had family members who had this learning style, but as she illustrated, so do many successful people:

> And he's (her partner) dyslexic so it's interesting I did wonder, what do they do about people who have learning disabilities? It's kind of nuts, isn't it because our youngest is never ever going to be able to pass this test. He still struggles to get letters the right way round. And in all honesty the founder of this company was dyslexic. He's no longer the CTO but he was dyslexic, and he founded a multi-billion dollar company. You know it's often associated with entrepreneurial skills.

6.2 Economic consequences

6.2.1 The lucrative nature of testing and test preparation

From the test takers' perspectives, language testing had beneficial consequences because it allowed them to migrate and gain employment in an English-speaking country. However, the journey to gaining the required test score result for employment, professional registration or migration was often long, expensive and without the guarantee of success. The librarian from Poland explained that due to the costs of IELTS, which she had taken seven or more times, she decided to stop considering teacher registration in New Zealand. This is despite the fact that she had a Master's degree in teaching physical education from her native country. The university student from Singapore (M SNNS) had also taken IELTS multiple times. Yet, he had completed an undergraduate degree in Australia and was taking another in Health Sciences in New Zealand. For him the cost of taking the test was unfair "It's not about the grades, it's about the cost because for me personally, I know that my English standard is good so when I get a 7.0, I look at it and go pfff … I just have to retake it."

For the two test takers from the Philippines, the testing system had gained the reputation of being a money-making business. The Filipino medical scientist (F FNNS) stated that:

> I think people are thinking it's becoming a money-making business. Yeah, an avenue to make money and that people are just ripping off so they keep failing participants, just for them to make money. That's the view now.

This participant had gained employment commensurate with her qualifications, so the test had beneficial outcomes for her. For the manufacturing employee from the Philippines (M FNNS), however, the possibility of continuing employment and permanent residence in New Zealand was contingent on his test score results. He explained that he could ill afford all the related costs because if all the preparation costs were taken into account, the overall cost could be $1,000.

> It costs us a lot. To pass the IELTS exam, you need to go to coaching lesson, which you need to pay for it. At the same time, we have to skip sometimes in our job, just to get to coaching lessons and also the time we go to take the exam, we need to refresh, to relax at least and skip overtime so it costs about a grand.

The primary school teacher from Germany (F GNNS) took a wider perspective, commenting on the testing system as a business:

> And the money that goes into this. There are 75 students sitting there all paying. That is a huge amount of money. You know on my second test I sort of talked to a lot of people taking the test with me and literally 99 percent were doing it the second or third time around.

This participant was also critical of the fact that there was very little feedback about her test results and if she wanted feedback, she had to pay for it: "And then IELTS charges you if you want more detailed feedback. You can do an on-line test and you can send it away, but they charge you for that and I sort of felt that they've already taken a lot of my money" (F GNNS).

Not only non-native speakers but also native speaker test takers and employers noted that the tests for migration are likely to be money-making entities. The native speaker primary school teacher (F1) looked at the need to prepare for the test as becoming, in all likelihood, an industry for native speaker and non-native speaker test takers alike:

> Oh, I think most New Zealanders wouldn't pass that, not without coaching beforehand being taught to the test because I don't think this is … yeah, it's like anything it's going to generate a whole industry around people having to pay to go to classes to pass this particular test whereas it's not going to assess general social competence or the ability to operate successfully in our society.

Similarly, the ESOL employer (Emp 4) worked with students who were migrants to New Zealand and explained that there were a couple of ways for migrants to demonstrate their language skills but that each pathway was expensive:

> You know some of the spouses we get of people who have already got their residence, they're not going to need this. I mean they're not the primary applicant and I think they need to get 6.0 so that's a lot of money trying to get them through the IELTS but if they don't get through the IELTS, it's a lot of money they have to pre-pay for English tuition.

The retail director (Emp 9) questioned the expense of the test, asking:" Is there a fee for this test? Because if they're doing it multiple times it's not realistic. It's quite an expense." Similarly, the waitress (F3) questioned the cost of a high-stakes language test, which she perceived as being unfair for migrants:

> So, does it cost? So, they're paying, they're paying a lot of money to basically be told you've failed, go home. It's a waste of everyone's time, it's a waste of money. It's like a scam – just to get money out of immigrants. It's stupid.

The ICT manager (F2) and retail assistant (F3) suggested that associated costs were not merely financial but also psychological and social because: "as long as you are in that sort of range, there's some sort of leeway and it's just a humane thing to do because the human cost of failing is huge." (F2). The retail assistant (F3) commented on the need for Early Childhood teachers to score 7.0 in IELTS by remarking that this was illogical:

> I think something's not adding up. I don't understand that. I know that early childhood teachers have to do a lot of work for each child. They have to write assessments and things. But really that's pretty ridiculous - there's no humanitarian or something practical, there's something missing. It's like logic or sensibility has been taken away.

In fact, the German primary school teacher (F GNNS) had experienced this anguish when she had found it difficult to gain 7.0 in her writing "To me, it meant having a registration number or not having a registration number. It meant the world – and for half a point."

6.2.2 Business economic consequences

Migration policies, including the need for migrants to take language tests, appeared to be a financial burden for employers in the business, IT, manufacturing and retail sectors. This was because many businesses could not survive without employing migrants and these migrants had to take a language test to remain in New Zealand. The manufacturing employer (Emp 8) described his frustration and the costs of trying to retain migrants the company had trained:

> It's devastating for us. We had a situation earlier in the year where we had someone who came here to study and studied the topic that you know is of interest for us. We employed them on a student visa to work for a period of time and, you know, you train them, you develop them, they know your workplace and because we had previously tried to recruit that externally with no luck. It came to their visa and immigration said 'no', they had to go and we really had to fight for her, to try and keep her here because it wasn't just on her behalf but from a business point of view continuity is quite important as well as just the amount of cost and effort in hiring and re-training and getting someone to the point where they are competent at what they do and so it makes it very, very difficult for us. So, we have had to say good-bye to people.

The food and safety consultant (Emp 10) was also concerned about the cost of training: "What tends to go wrong is every time you change staff and you train because you know the job so well, it actually takes a year or so for them to be there. But you know you haven't got that much time. It's

costly to train people up." The retail employer (Emp 9) employed almost exclusively migrants because "Years back we used to have quite a few Kiwis, but it's just changed so much, you can't get them. Kiwis now want the 9-5 jobs and their weekends off and the lifestyle and you can understand that but as an employer it's a problem". However, he noted that employing migrants who could not necessarily stay in the country meant a constant cycle of training staff.

6.3 Societal consequences

6.3.1 The role of migrants in New Zealand society

Many employers claimed that their businesses relied on a migrant workforce, who needed to gain 6.5 IELTS for a work visa. Yet, the political environment had, according to the retail and manufacturing employers, changed. New immigration policies were becoming so complicated that they could not retain valuable staff. The manufacturing employer (Emp 8) described the situation for his company:

> We employ 2,300 and probably 15 – 1800 of them are blue-collar workers and we just struggle to attract New Zealanders, there just aren't that many of them out there and probably the ones that are left, often aren't the ones who want to work, so we do rely heavily on migrants and what we're finding at the moment, and even though we are a credited employer, is the amount of scrutiny that's being applied to the applications that go through for us are significant. People are having to provide ridiculous amounts of information to justify their existence as to why they are here, regardless of how long they've been here, or regardless of what country they're from either, so I think there's just a certain change in immigration that's making life very, very difficult for migrants.

This employer (Emp 8) expressed his dismay with the immigration policies:

> It's so disappointing when you've got skilled talented people who can find employment but who are unable to stay here because they can't pass a writing test, it just blows my mind. And Kiwis may struggle with this test, so I just think it's ridiculous ... It's devastating, and I think it's to New Zealand's detriment to be honest (Emp 8).

The situation was more difficult for the retail employer (Emp 9) because retail was not included in the essential skills list by Immigration New Zealand. Yet, in places like Queenstown employers relied on migrants:

> Well for retail we're non-essential so that's why we struggle to get some permit ... I mean every employer down there must be doing the same, providing some application for a

> person who is living in Queenstown where no New Zealander ever applies for a job in retail or hospitality. It's like in Britain where everyone who works at the coalface where tourists meet is from Poland or Eastern European countries or backpackers - New Zealanders, Australians, everyone, and we're the same now in Queenstown. The place survives on those form of workers.

It seems that the majority of employees in the food manufacturing businesses are migrants. As the health and safety consultant (Emp 10) explained **"**I do deal with Kiwis in the brewing guild. In the bakeries, there are still Kiwi owners but it's just getting people to work, to work the hours and to do that type of work." For this employer, a score of IELTS 4.0 would be relevant in her sector and, as she recognised migrants came to New Zealand expecting to work hard "The reason why they came here is for a better life and a better life will mean they have to learn the language but not to this extent ... If everyone has to pass this test, you'll skew the workplace and you won't get the workers."

6.3.2 Education and the monolingual mindset

Paradoxically, participants lamented the lack of attention to linguistic accuracy by teachers in the New Zealand education system, and yet, society still appears to expect high levels of proficiency from non-native speaker migrants. The policy of demanding high IELTs scores of non-native English speaking immigrants, when high standards of proficiency are not expected in the wider society, can be seen as being discriminatory. The IT employer (Emp 5) did not believe that school students were taught English in a systematic way:

> If we compare this with the kids nowadays. I don't really feel that the kids are taught formal writing to the extent that we were and you know certainly my kids are going to school and they are doing a lot of copying and pasting stuff off websites so when they're trying to get the word count they are copying and pasting more sentences and then maybe revising them in their own words so they're not actually plagiarising.

A similar concern was voiced by the accounts manager (F5). She questioned how many policy makers would pass this type of test and what that implied about the education system in general:

> Yeah, I wonder how many of them would pass. No, I think it says something about ... well if that's the level of education that we're all aspiring to, then it says that something's wrong with our education system because we've all failed. I think we're just asking people to do something that's ridiculous. If we can't do it, why should we expect people to ….

While a certain proportion of New Zealand society may not have high literacy skills, native speakers often expect migrants to be proficient in English (the monolingual mindset). As the librarian from Poland (F PNNS) suggested in 5.4.3, it is often considered to be alright for native speakers to make mistakes, but not for non-native speakers of English. The university lecturer (M1) noted, if non-native students were unable to communicate effectively, it was often presumed that they lacked intelligence "I think one of the real problems though is that when someone can't communicate, the audience assumes that they are not intelligent."

The issue of intolerance towards accents was commented on by employers such as the IT employer (Emp 5):

> Customers tend to have a very low tolerance of people with accents … So we are very conscious that there are low tolerance levels or perceived communication issues because customers are ringing up and are already tense because their application isn't working and something won't turn on or turn off so they are already frustrated and if they have to talk to somebody, who they don't feel understands them immediately, they are going to lose the plot.

The ESOL employer (Emp 4) referred to the issue of people accepting the accents of students she worked with: "Their English was good, but it's often accents because Kiwis have real problems with accents. It's getting better, but it is an issue for a lot of people"

While the senior educator (Emp 2) was aware that New Zealand students complained about teachers from countries other than New Zealand having an accent, she had counselled them to listen more carefully:

> I've had a range of experiences where students say I don't really understand that teacher and I've said, you actually just have to listen because their communication is probably better than your average Kiwi teacher and their understanding of language and the way that they use language, but you have to stop and do the listening.

The necessity for New Zealanders to make an effort to try and understand people from other countries was referred to by the ICT manager (F2). She pointed out that today's workplace and many societies are now multicultural so there is an onus on all nationalities to endeavour to understand each other:

> And also what about English speaking doctors and nurses, you know who are dealing with people from the Islands and foreigners who come over from ... their patients, you know why

> are you only expecting people to do the English thing, you know, surely doctors and nurses who deal with foreign patients should be expected to have some knowledge or some way of communicating, it works both ways, because we are multicultural now, everywhere there's this big melting pot, it doesn't matter where you go.

Racial issues were discussed by the primary school teacher (F1) in a stronger way. She suggested that immigration and language policies may lead to more overt racial discrimination: "We're going to end up with the white is right thing by running this because that is what Australia has done."

6.3.3 Summary

New Zealand needs migrants with different educational and professional levels to work in a variety of employment sectors. Yet, for these research participants the IELTS test appeared to be testing language proficiency at a higher level than was necessary for workplace communication. The participants also remarked that people with dyslexia who may not do well on traditional tests can be effective workers. Language literacy levels among the native speaker population is not necessarily high; yet, non-native speakers of English are expected to score well, and this could result in racial discrimination. Language testing was regarded as being a lucrative enterprise with psychological as well as financial costs. Economically, language assessment policies can lead to a lack of business continuity according to the employers in this research.

6.4 The consequences of policies

After viewing sample language tests, test responses and the test score results of native and non-native speakers of English, all of the participants in this research suggested that there was probably a misunderstanding of the function of the test, the test content and the meaning of score results by decision makers (test users) at the policy level.

6.4.1 Lack of language assessment knowledge

Firstly, the lack of understanding of what is involved in a high-stakes international language test was mentioned by some non-native speakers. Talking about the registration bodies and the immigration authorities, the Malaysian PhD student (F MNNS) would like to see more awareness of what these high stakes language mean because she seemed to believe it was easy for authorities to rely on the test developers:

> I think they have put that in place, to have a standardised test. It's like a benchmark kind of thing. But I don't think they understand it. They have just left it to the testing body to come up with the best possible way to measure it and they have just taken it as it is.

This lack of language assessment knowledge was also mentioned by the majority of native speakers (test takers and test users). Looking at the need for migrants to New Zealand to gain a score of 6.5 IELTS in all four skills, the native speakers who had taken a sample test recognised that the immigration authorities probably did not understand the test or the test score results. The university undergraduate student (M5) commented thus " I think that people who choose the numbers don't understand the test and if they do understand the test, and if they had sat through it and seen what they'd got, I have no idea why they would put it at 6.5 for entry to NZ."

The university lecturer (M1) argued that policy decision makers had possibly not taken enough time to look at the test in depth and whether the test was fit for purpose:

> So, there is a misunderstanding about what the test is doing or what it means or whatever. I'm not sure if using the band scores is any worse than the test as a whole - it is all flawed in the same way ... Again, it's what is fit for purpose. Obviously, immigration NZ haven't thought about it at all in terms of what is fit for purpose. Clearly if the same band scores are needed for a sheep shearer and a university lecturer, they haven't thought about it.

Likewise, the primary school principal (Emp 3) reiterated the need for decision makers to evaluate if this sort of test is fit for purpose:

> Well, obviously immigration don't know what these numbers mean, and I get that to a certain degree, but you'd hope that the top people at immigration would. Is it fit for purpose? Why do we test? And is that test what people are going to achieve? ... you're looking for a new life change at 35 or 40 and you have to do this? I don't see the fairness in that. To me there's more important things than being able to write, you know writing is a part of it but being able to hold a conversation, absolutely.

The question of whether decision makers understood the difference between communicative competence and what is being tested by IELTS was raised by the Master's level student (F6):

> I think what the policy makers are trying to attain is - can they communicate in English, can they, do they have a level of English that is high enough that they can be understood and understand? This test doesn't necessarily do that.

As the ESOL employer (Emp 4) argued, the test score results did not relate to a person's workplace competence and that migrants who scored 6.0 or less could actually be an asset for New Zealand. Here, she is describing migrants such as tradespeople who have been brought in to work in New

Zealand for the Canterbury rebuild and who often have specific industry skills but for whom the level of 6.0 IELTS is often a ceiling level which they cannot transcend.

> But then I don't find that surprising because I've worked for government for such a long time, and that's the way things work, it's all very arbitrary. They like boxes that they can fit people in - they don't actually relate to the people (Emp 4).

As the telecom technician (M2) pointed out: "I think they'd be horrified at who they are excluding based on what the test actually measures."

Some of the participants proffered reasons for the lack of language assessment knowledge amongst decision makers. The ESOL teacher (M4) had previously worked in government, so he recognised that budget cuts and time limits could lead to a lack of awareness of the implications of using this type of test as a political boundary object:

> Because the people who are marking it or the way in which this exam is being applied, it is being mis-applied. There's too much of a disconnect between the real world and what this was originally designed for or alternatively, it is because I don't have time to prepare exams and all the rest of it, so this is a ready-made system, it's easy because it's got numbers and as long as I've got numbers I think ok I have to have a 7 and a 6 and I'm mis-using it completely from what it was originally designed for but because I haven't got time, I've got budget cuts but I've got to use something because I've got to up the barriers so I can use this instead of that.

The reason for this lack of understanding was connected with authorities, who ask migrants to gain an IELTS score of 6.5, being ill-advised. The manufacturing employer claimed (Emp 8): "I really, truly believe they don't understand the implications of what they are actually asking. I think that maybe they've just got some bad advice around what should be a minimum level."

The issue of whether policy makers understood the function of language testing in light of the changing nature of literacy in many parts of society was mentioned by the primary school principal (Emp 3). He could not understand why the native speaker lecturer had been given a mark (6.5) lower than that required by overseas teachers to teach in New Zealand (7.0) and queried if the test was being adapted to changes in society:

> Well, the academic professor who's doing writing, setting tasks for students and has probably got a doctorate but he only got a 6.5. It seems is the test actually relevant?... Do I

> think it's fair? Not necessarily. And it would be interesting to know, when was the test set, has it been the same test for a long time? Society has changed.

6.4.2 Need for language assessment literacy

The need for decision makers such as Immigration New Zealand and various professional registration bodies to have a better understanding of international language testing systems was perceived to be important by the majority of my participants. This was because the required band score results of 6.5 for immigration and 7.0 for professional registration in the reading and particularly in the writing sections, were deemed by test takers and test users alike, to be unrealistic. The ESOL teacher (M4) explained that it was important for decision makers to gain in-depth knowledge about the implications rather than following standards set by other institutions:

> I think decision makers, people like the Education Council are being completely unrealistic ... I think educators are amongst the worst educators, and those who set these standards need to think very carefully about why they are doing it. I think that perhaps they are clutching at straws because there's a standard and someone else has set it and there's an assumption that this must be a good standard and so we'll adopt it without thinking through what is it actually that we would like – is it attitude, is it adaptability, all those sorts of things?

The business consultant (Emp 6) suggested that employers in New Zealand need more education on what this type of language test and the related policies mean for their businesses:

> That would be my concern is that it's only one measure but it's only because I understand the IELTS that I would take it with a grain of salt. So, employers really don't understand it. There's not enough out there and that's possibly where it could improve. It's getting some education out to employers. You know a 7.0 does not necessarily indicate that this person is fabulous in all areas of English or you could look at somebody with less than a 7.0 but they don't get the residence.

This participant (Emp 6) commented that one of the issues is that IELTS results do not correspond with workplace competence, but that education is required so that there is less miscomprehension. In response to my question regarding whether test score results lead to workplace competence, she responded:

> I don't see any correlation at all actually and the fact that employers wouldn't understand them anyway, just adds to it. So yes, I think there's absolutely a case for lowering that IELTS

level. You could get people up in arms but the people who go up in arms don't understand IELTS.

6.4.3 How to introduce language assessment literacy

Many of my participants mentioned the arbitrary nature of picking a number as a cut-off score on a language test and trying to ensure every prospective migrant fit that number in a Procrustean bed way, as described by Spolsky (2008). Many participants mentioned that this did not appear to be fair for test takers, or beneficial for New Zealand society.

One recommendation made by test takers and employers was for wider consultation between test developers, language assessment researchers and test users. As the manufacturing employer (Emp 8) proposed:

> Well I guess they have to create some form of criteria, but I think it's got to be realistic. So, I think you just can't pluck an arbitrary figure and say everyone has to hit that target. I think there needs to be some consultation with the industry or with academia to say what's actually reasonable, what's fair, um a fair score.

A second recommendation was to assess the scores that the average New Zealander would gain on a test such as IELTS and to set score requirements according to these scores:

> But I don't know who decided it should be set at 7.0 but what they should do is if they want to set a level, they should go and take a hundred people off the street, test them and then whatever the average is for the people who are successful, that should be the level you have to meet not some weird arbitrary you know… that's just patently ridiculous (Primary school teacher, F1).

This viewpoint was echoed by the accounts manager (F5):

> This is pitched way above ... so why aren't we pitching it at an average New Zealander level, why are we making it so much harder, cos this isn't a 15-year old level, is it? … That was my thing before - if the national average of whatever our literacy level is 15 or 16, that's what we should be testing for. I guess that it would be nicer if the people who are coming into a country are improving our intelligence overall but it's not a need.

Some participants, such as the telecom technician (M2) went further suggesting the need for the test developers to collaborate with test users to ascertain the average score that would be likely gained by native speakers in different occupational groups:

> Well, maybe the test should be normed by reference to the population groups towards which it is intended so if you're a … it seems silly to request a score that's higher than what a native speaker would have so maybe they should set the level of entry into NZ based on what the average native speaker gets and maybe if you want to be an electrician, maybe IELTS or the government should be putting in the money into say what level of score is required for a particular occupational group.

Some non-native speaker test takers commented that testing the average native speaker in different professional groups would lead to more reliable and fair outcomes when using language test results. The librarian from Poland (F PNNS) suggested:

> And I would say just go and do research about Kiwi primary school teachers and what they would score and if they score like on average 6.5 then international primary school teachers, who is non-Kiwi, is only supposed to get 6.5. Or, if the average vet would score 6.0, maybe that's supposed to be the same benchmark for the non-native speakers. So yeah, be more specific about that kind of jobs.

The third recommendation was for test users such as politicians or policy makers to sit the test themselves to ascertain their IELTS level, and their perceptions of the relevance of the test. The waitress (F4) put it this way:

> I'd just tell them to chill out. They just need to relax a wee bit – yeah, they should take the test themselves and see if they can do it – it's hard. They need to do it themselves and see what it feels like because if they fail, imagine the disappointment they feel and they're a native English speaker.

The retail employee (M3) and the university student (M5) agreed, stating "I'd probably say take the test yourself and then ask yourself if you think someone who can't speak your language can pass it".

The Chinese businessman (M CNNS) stressed that it may help immigration staff members to take a sample test because an increase in language assessment literacy may lead to more robust decision making:

> I think it's unfair. I should let the immigration officer should sit the test. If they can't get higher than me why should I not be able to immigrate to NZ. I think people think 7.0 is an indicator of how good your English is but actually people could, in some areas their English is really good, for example an engineer could read a structural report, no problem, but if you

> ask them to read a newspaper it probably would be more difficult. But the immigration thinks if they get 7.0, they can read anything.

For the ICT manager (F2), it was important for decision makers not just to experience the test themselves but to do so on a regular basis:

> I think people who make decisions they should experience this often. You know if you set these things, then you need to be able to, every so often, re-sit them and trial them constantly. You know every year you should be made to sit this thing. It's like anything in industry, like if I ask my team to do something, I have to make damn sure that I can do it as well and I like have to put my toe in the water every so often myself so that when they say to me there are these issues then I have actually experienced it myself. And that's what I think decision makers should do, they should dip their toe in the water regularly and often.

Finally, the issue of fairness and discrimination were discussed. The primary school teacher (F1) and the university lecturer (M GNNS) both suggested that every teacher in New Zealand (native and non-native speakers) should be required to take the test in the interest of fairness and equitability. The primary school teacher stated: "So teachers should but at the same time every teacher that goes through a New Zealand teacher's college or teaching degree should pass this as well even if they were born here." The lecturer from Germany (M GNNS) had the same opinion, commenting on the issue of equity:

> And, it would be interesting to see how native speakers would be able to do that unpacking and response to these texts. Um, from a living perspective, and now this is the academic speaking, thinking about democratic citizenship and contribution to society, yes, I think it would be actually quite good if everybody would be able to unpack these things. I think the challenge that I would see is the challenge of equity or between expectations of native speakers and expectations of non-native speakers ...We have a maths test here in New Zealand now so that teachers need to show that they have certain maths skills, so I think they should show English writing skills as well… I think that for teachers the requirement should be across the board so for native and non-native speakers to make it equitable….If there's a reason why we expect 7.0 in writing for teachers, for example, and the reason is most likely that we want to have people who teach our young generation to have good literacy skills , then we should require this from everybody.

This non-native lecturer (M GNNS) recognised that many native speakers of English may overestimate their English abilities and that this leads to discrimination in the workplace:

> I think there shouldn't be an arbitrary number, but every industry needs to look into a couple of real time tests and to do this job that we want to have in this industry, we would expect everybody to write at this or this level. It also wouldn't hurt for these people to take the test themselves to compare their own outcome because we, as human beings, are very easy to over-judge our own abilities and our own outcomes, just as a way of reaffirming our own personality and self-esteem so for everybody who makes such a high category judgement to go through the hoops themselves to see where they would sit and to manage their own expectations well would be sound advice, I would think. You know proficient writing is hardly ever produced in normal life so there is an element of discrimination here.

6.5 Possible solutions

6.5.1 More realism and flexibility

Taking the employment situation and the literacy levels of native speakers of English into account, the native speaker research participants recognised the lack of fairness in using this test for migration purposes and ultimately called for more flexibility in the use of the test score results. There was particular concern with the writing section of the exam.

Some native speaker test takers suggested that test users should be more realistic and take lower test score results as proof of communication skills in English. The retail employee (M3) said "Either change the test to easier or lower their expectation of the score, a little bit. Like, make it … because they have to score above our scores on average so make their average slightly below what our one is because we obviously speak the language", while the ICT manager (F2) thought "I think they should just make it easier, like make it more relevant." The notion of making the test more relevant to society was shared by the telecom technician (M2):

> Um, the people who make the decisions further up the chain, they need to become aware of how difficult it is and how in many cases how irrelevant it is. Well they're far too difficult for the type of people they're bringing into the country.

The Chinese businessman (M CNNS) and the university student (M5) both suggested having lower test scores for prospective immigrants at the beginning of the immigration process to allow people living and working in the country time to develop their English language skills. As the Chinese NNS (M CNNS) pointed out:

> I think people need 5.0 for immigration and an interview would be fair for people who want to immigrate to New Zealand because English is a progressive learning curve so if you give

> me the chance, I get work and maybe you can test me in two-years' time. Some fast learners, even if their English is not very good, they can understand each other, and they can pick up things but just judging people by IELTS is not fair, it is just a number.

The university student (M5) commented on the same topic because he was under the impression that progressive testing was undertaken in Germany:

> Doesn't Germany have a system where it's easier to get in but over time you have to improve or something? I've heard that from somebody so to get citizenship or residency you have to do a test which is fairly easy but after a year or so, you have to re-sit it and have improved a certain amount. I kind of feel that might be a better alternative.

The only person to comment on the time limit that policy makers allocate for test score results was the scientist from Malaysia (F MNNS) who noted, like the two participants above, that when language is used in context, it should improve rather than worsen.

> It's worse when you have to renew it every two years, so it's only got a life span of two years, so you have to re-do it even though you got a score of 8.0. You know, it should be a test that is done once and that's it. If you need to measure it, you can't lose a language, you can't be an 8.0 and then suddenly be a 5.0 in the next two years. It should improve in a sense, so I think that's not right. I think it should be just a one-off thing.

The non-native speaker test takers were often more positive towards the IELTS test with the production manager from India saying it was essential for all migrants to gain an IELTS score of around 6.0-6.5 but that an overall score rather than minimum levels should be acceptable*:*

> I would like to say something because 6.5 for residency, it should be compulsory, I think. Apart from that, not only a student visa. I think those who come on a spouse visa should pass IELTS with at least 6.0… it's for their own benefit, so that they have their own skill built up for English, they can easily find job over here, so that they can communicate with their team members ... For student visa, or spouse visa or any kind of visa, I don't think they should set in each criteria, they should set overall (M INNS).

Some native speaker test takers thought it best not to use a specific number as a boundary point but rather to accept scores within a range, and to have the flexibility of offsetting a low score in one section with a high score in another:

> I think that maybe instead of requiring the competency level in each separate category, maybe do an overall average and then having a band instead of a minimum cut-off and

> allowing a little bit of flexibility for the employment organisation to say well look you've got an average score of 6.5 or 7.0 and one area can subsidise another a little bit … Or they average it out so if you get say a 9.0 in listening and a 6.5 in reading it's averaged. So, they need to let someone in who may not necessarily be good at one thing but is excellent in another. We've all got our strengths and weaknesses *(*Telecom technician*,* M2).

Similarly, the university student from Singapore (M SNNS) had experienced having jagged scores and argued that the scores for each skill should be flexible, depending on the area of employment, especially since writing may not be a necessary skill in some occupations:

> But the average score doesn't really show what your individual strengths are because I remember getting a 9.0 once but I was pulled down because I had a 7.0 so when you look at the total aggregate you see that on average this guy is an 8.0 but when you look at the individual scores you can see that he's very, very strong at speaking and for certain jobs, like retail, you need people who can speak well. It's just that in certain jobs you don't need writing so why bother about writing.

The primary school teacher from Germany (F GNNS) explained that the Education Council in New Zealand had been very understanding about her inability to score 7.0 in the writing module, but that they needed to be more flexible because this score may be a difficult score for New Zealanders to attain and therefore unattainable for many non-native speakers of English:

> Well, the Education Council say they have an IELTS panel and the panel has approved my scores. I was in contact with them a lot and it didn't feel that they didn't understand or that they were being rude, but they still have to do things by the book I guess like in their user manual. I didn't feel like they didn't care, well not my case worker anyway, she really did care. I feel there should be a face to face communication with immigration, anyway. Maybe the easiest and fastest way of looking at it is lowering that band score for the writing by a point, from a 7.0 to a 6.0 because there is still some very competent writing going on and 7.0 seems unreachable almost.

Likewise, the doctoral student from Malaysia (F MNNS) suggested that a lower score for the writing module should be acceptable because of her perception that this section was more difficult than the other three skills. For her (F MNNS) the writing tasks were particularly difficult:

> Especially when it comes to written work, it takes years of work in terms of grammar, punctuations and all those things and I feel it brings down the other scores and sometimes

> it's not fair…I think you need to reduce this one for writing in terms of … because it's a very complex task compared to other scores and you cannot measure it the same way you measure the other skills. So, I think if you do well in all the other skills and your writing is a little bit lower, I think you should be considered as a competent speaker anyway.

While the ESOL teacher (M4) agreed with the political objective of using language tests as boundary objects to restrict immigration, he felt that this was this should not apply to the writing section or that the prerequisites for the writing section should be lower.

> So, I think there are outside factors that influence this, so why do we demand a particular standard and I think this gets tied up with politics about the ease in which people come into our country and I think that's good. I think we should restrict immigration but sometimes for the writing thing, even I wouldn't get into whatever it is.

The employers seemed to agree that the usefulness of high scores in all four skills was not important in all occupations. As the business consultant (Emp 6) explained:

> Maybe, according to an employer, they might say 'Look, we don't need them to be proficient in reading' and, in fact, most of our Kiwis – we know how low the literacy rate among Kiwis is. Most Kiwis wouldn't get that so why are we forcing our migrant workers to get it? So, that's been a conversation now and then with employers.

According to the retail employer (Emp 9), the lack of flexibility in governmental policies was detrimental for his business because they did not need to employ highly educated people. For this reason, employers wanted policies which include language test results to be more realistic for society as a whole:

> Well, if we're just looking at our business, they're not really relevant to a point because we don't need those skills but … they're almost non-achievable for a person who would work in a retail store. My argument with immigration would be that there's different levels of education and intelligence needed for any job and when there are job shortages we should be a bit more lenient here because we need them, and we can change it later. I suppose it's being more flexible.

Furthermore, the retail employer (Emp 9) could envisage this lack of flexibility having a negative impact on the New Zealand economy in the future:

> It's sad really, it's a waste … An application for residency from an interview process, maybe with the employers and then how good's this person for our society, more of that sort of

> expectation and lowering this level of English a bit because we've seen the results and there's some very intelligent people and they've just scraped through really, haven't they, if you look at those results. I don't think it's right, I don't think it caters for … where NZ is going to have issues with is IT because that's one of the quickest growing areas in the global economy. So, I think some of those areas there need to be more weighted for what's going to be good for the country going forward rather than just a one-off type test.

6.5.2 The role of policy makers

The role of policy makers and the issue of who should make the decisions about a person's language was commented on by the ICT manager (F2)

> Yeah, that would mean devolving, like if the Education Council sets the cut-offs instead of having a minimum cut-off you could say you have to be an average score in this range and allow a bit of flexibility and allow the final decision to be made by people who are sitting across the table from them instead of some faceless bureaucrat in Wellington that sets the cut off.

Similarly, the university lecturer (M1) pondered if decisions about migrants' communicative skills were being made by the right authorities:

> Well, communication is important obviously but then we should be aware that this isn't perfect and that we should start at a maybe lower level and then add on. It's not difficult to assess somebody's communications skills. So, we are ceding the responsibility of assessing people's communication skills to some unfaced individual in Australia ... You know if you walk into a bar and say hi to someone and you can tell within a- few minutes if they can communicate – to some level. And then you can look at a piece of writing and assess not whether they can use the exact words but whether they can communicate whatever.

The need for consultation between stakeholders and a more in-depth examination of the test and its' purpose in society was mentioned as being necessary. As the ESOL teacher (M4) explained, the role of the test has changed with it now being increasingly used for business and migration purposes as opposed to university entrance, but the implications of this widened application need investigation:

> So, it's borne out of, because designing something with finesse takes more time and skill and we're now applying it in so many different areas that we've actually lost the reason why we originally did it … Well, there are two types of decision makers. I think that those running the

> IELTS stuff ought to make it a bit more clear as to the criteria and the standards and the political decision makers ought to be a bit more refined in the way that they apply these standards ... I think they need to start talking with the people on the ground.

6.5.3 Introducing a practical New Zealand based test

A final suggestion made by a number of participants was to introduce a more practical, New Zealand specific test. As the telecom technician surmised:

> Well maybe there's a bigger question, that IELTS is not relevant any more maybe the people that run IELTS are protecting themselves and making it very, very difficult. Maybe there's another test that needs to be developed, something that's more practical (M2).

The Chinese businessman (M CNNS) had studied courses to prepare for the IELTS test but had also taken a New Zealand lifestyle course, which he had found much more practical and useful for living in New Zealand "I think every immigrant should take a New Zealand lifestyle class, you know. That's more important because that's why they are staying in New Zealand they've got to learn how to survive."

The business consultant (Emp 6) had long considered it necessary to design a test specifically for the New Zealand workplace. As she argues, IELTS is a well-respected international test but a more practical, New Zealand competence-based test may be more applicable for the New Zealand environment.

> Because it is an international test and we can respect that, and because it's accepted everywhere that's great and a lot of our workers have worked all over the world so for them it's a good base line thing, but I do think when they come in we possibly need a New Zealand based workplace English test ... Certainly, I would love to see a work-based competency test but that's tricky too because it is different across different industries but a New Zealand one with New Zealand accents, New Zealand slang and writing in a more practical, not so academic way.

6.6 Summary

The participants in this research highlighted that test users (New Zealand employers) need language assessment literacy (LAL) to understand the content of language tests, their relevance for the New Zealand workplace, what test scores signify and who these tests may be excluding. Language tests also need to reflect changing trends in writing for the workplace. There is, therefore, a need for consultation between test developers, test users and language assessment researchers. For

language testing systems to be regarded as being ethical, there needs to be more realism and flexibility in what is acceptable for migration and professional registration purposes. This may entail including industry insiders in setting cut-off scores which are relevant and appropriate for their sector. Other possibilities include the introduction of an Australasian test of communicative competence, or a language test for different professions as described by Knoch and Macqueen (2020).

CHAPTER 7

DISCUSSION, IMPLICATIONS AND CONCLUSION

7.0 Introduction

This thesis explored and analysed the perceptions of test takers (native and non-native speakers of English) as well as test users (English speaking employers) regarding the use of high-stakes language testing for migration, employment and professional registration in New Zealand. This is because there has been a paucity of research on the perceptions of native speaker test takers, non-native speaker test takers and employers regarding the test content and constructs being measured in English language tests compared with the language used in the contemporary workplace. This chapter uses the research participants' perceptions to discuss the validity (through domain and extrapolation inferences) and ethics (through the decision and consequences inferences) of using language tests for employment, professional registration and immigration purposes in New Zealand.

Chapter 4 (findings part one) documented the participants' reflections on a personal or individual level to answer RQ1: How well do test takers and test users perceive the relevance of the assessment and assessment tasks to the New Zealand domain (domain inference)? Key findings included their viewpoints that IELTS is a meaningful and relevant language test for New Zealand, but one that is not understood in depth by test users. All stakeholders appreciated the need for communicative competence in the workplace, and for this reason strong speaking and listening skills were valued. The conclusion reached by all the research participants was that this type of language test was assessing educational level or higher language proficiency (HLP) as opposed to communicative competence. The second part of this chapter presented the participants' views regarding their ability to use and interpret test scores and their feelings towards the required minimum standards for the New Zealand context. These viewpoints were used to provide data for RQ2: How well do test takers and test users feel they can interpret and use test score results (decision inference)? The results showed that test takers and test users could understand and appreciated the cut-off scores required for the speaking and listening parts of the IELTS test. However, they could not use, interpret and did not appreciate the need for high scores in the reading, but particularly in the writing section of the test.

The participants' views of how the English language is used at a group or workplace level were analysed in Chapter 5 to answer RQ3: How well do test takers and test users feel that the IELTS tasks and the constructs being tested reflect communicative language used in the New Zealand workplace (extrapolation inference)? It transpired from the native speaker participants' perspectives that high

standards of written accuracy and complex language are no longer expected, even in the educational sphere. A further finding was that all participants (native-speaker test takers, non-native speaker test takers and New Zealand employers) did not expect their work or study colleagues to gain high test score results in the reading and writing sections of this type of language test.

In Chapter 6, my research participants voiced their opinions about the likely social, economic, societal and political consequences of using language testing in New Zealand for migration purposes. These opinions provided data for RQ4: What do test takers and test users perceive to be the consequences of test use for the New Zealand context (consequences inference)? The economic consequences discussed included the costs of test taking and test preparation together with the costs for businesses. Social consequences mentioned by research participants included issues of discrimination against certain nationalities, those with learning disabilities or those working in non-skilled employment spheres. In particular, the native speaker participants in this research study felt that policy makers and professional bodies need greater language assessment literacy to understand the nature of language tests, such as IELTS, which are being used as part of migration and employment selection processes, in order to avoid unethical decision making.

While there may be merit in using international language tests such as IELTS and TOEFL, the findings reported in Chapter 6 indicate that the need for written workplace communication has changed since these high stakes language tests were introduced as benchmarking tools for immigration and professional registration policies. As Knoch and Macqueen (2016) pointed out:

> Communication in the workplace is changing rapidly and it is therefore important to constantly review, for instance the types of tasks included in an LSP [language for specific purposes] test to ensure that they remain current in terms of the TLU domain. Technology is transforming the way professionals communicate and this might affect the construct tested and approaches used to testing, including the test task" (p. 300).

A common question raised by participants in my study was: 'Is this test fit for purpose?'. The resulting research question was: To what extent are language test scores used in valid and ethical ways for the purposes of immigration, employment and professional registration in a New Zealand context?

In this final chapter, each of the research questions and the consequences and implications for New Zealand society will be discussed in sections 7.1 – 7.5. The research reported in my thesis is necessary because, as Ryan (2016) argued, there has been a dearth of research into the lived experiences of test takers or into the impacts of language tests at individual and societal levels.

7.1 RQ1: How well do test takers and test users perceive the relevance of the assessment and assessment tasks to the New Zealand domain (domain inference)?

7.1.1 High stakes language tests are relevant and meaningful for stakeholders

There has been a large increase in the use of high-stakes language tests in New Zealand. Although they are still being utilised by test takers and gatekeepers (test users) for academic entry, they are increasingly being used for immigration and professional registration purposes. The IELTS partners claim the following: "Established in 1989, and jointly owned by the British Council, IDP: IELTS Australia and Cambridge Assessment English, IELTS is now a household name in many countries around the world" (ielts, 2019). Therefore, the validity of this statement and the new purposes of the test will be analysed.

In this study, high-stakes language testing, and the name IELTS in particular, was meaningful for the research participants in the sense that it held significance for all of them (both native and non-native speakers of English); they were aware that these sorts of language tests existed, and the majority were aware of the acronym IELTS even if they had no connections with language testing or professional registration bodies. The participants who were in the education field and most in the business field encountered the acronym IELTS on a regular basis in their jobs. For native speaker test takers, the IELTS test was meaningful because it appeared to be similar to tests they had undertaken at secondary school level. Furthermore, participants noted that the international nature of the testing process rendered IELTS as being significant or meaningful for them. They also viewed the IELTS test as being objective, mainly because it is rated and administered by language testing professionals as opposed to immigration officials as the latter may have little knowledge of language testing. This means that many of this study's participants felt that IELTS can be relied upon. This finding aligns with the study conducted in Australia by Knoch and Macqueen (2016) who described how stakeholders valued aspects of tests such as test reliability and security. Likewise, the test was significant for test users in New Zealand, especially those involved with professional bodies such as the Education Council or Immigration Advisers' Council because test results are recognised by their counterparts in a variety of countries. This also corresponds with the research done by Merrifield into the impacts of using IELTS as an entry criterion by professional associations and registration entities in a variety of English-speaking countries (2008; 2011).

Hall (2010) claimed that IELTS has worldwide significance. Indeed, IELTS is widely used in New Zealand, and was a familiar acronym for most of the people I interviewed. However, only one participant had in-depth knowledge of the language testing process and that was the ESOL teacher

(M4). The native speaker test takers and test users who had not heard of the terms used for high-stakes language tests (IELTS or TOEFL) were mainly those in the retail sector. The majority of the employers in my study used the acronym IELTS, some of them quite regularly, without knowledge of what the initials stood for, but no one asked in the interviews what the acronym meant. This could mean, as Kunnan (2018) has argued, that test takers and test users accept that a test which is used as a boundary object or gatekeeping device (Macqueen, Pill & Knoch, 2016) is infallible because it has been designed by language testing professionals, and that test results convey a unidimensional 'truth' (Spolsky, 2008) about test takers, which should not or cannot be questioned. The unidimensional nature of language proficiency has been refuted by many researchers (McNamara & Roever, 2006; Pilcher & Richards, 2017; Ricento, 2014; Spolsky, 2008). All the test takers and test users in this study initially assumed that if policy makers such as immigration and professional registration bodies were using a high-stakes test, they had investigated it in depth. After viewing a sample test, their opinions changed as they decided that policy makers needed more language assessment literacy, as reported in 6.4.

In response to RQ1, it seems that IELTS has considerable relevance for stakeholders in New Zealand. However, the native speaker employers' understanding of language testing is superficial. For this reason, the test developers would be advised to spend more time and effort disseminating in-depth information about language testing to stakeholders in the sectors where they are promoting the test. These sectors include, amongst others, the education, business/IT and retail/manufacturing areas where many migrants are seeking employment and/or professional registration in New Zealand.

7.1.2 Test takers and test users are aware of/appreciate the skills being tested in the assessment tasks (test content)

In general, the native speakers (test-takers and employers) had no preconceptions about what a language test would be assessing. After viewing sample tests, they recognised the need to test the four skills of reading, writing, speaking and listening. Out of the four skills, the oral skills (listening and speaking) were regarded as being the most important for the New Zealand workplace for all my research participants. This could be because they represent communicative competence, that is being competent enough linguistically to understand and respond appropriately to superiors, colleagues, customers and students (in an educational context). Essentially, all participants in this research expected that a language test for migration should assess a person's ability to take part in everyday dialogues. In this respect, the IELTS test is valued because the speaking part of the test is conducted with a human rater, rather than with a computer as is the case with iBT TOEFL and PTE.

The native speaker participants (both test takers and employers) expected the reading test to contain reading material related to New Zealand life and work. The type of content expected was newspaper articles, work-related information and short texts on life in New Zealand. IELTS texts which are generally complex texts on the humanities, social sciences, scientific or technical subjects were not appreciated as being useful for life in New Zealand. Likewise, British based texts were not considered useful by the native speaker participants; many laughed when presented with texts about canals or the English countryside because the content was irrelevant to their lives. Half of the native speaker participants (test takers and employers) felt that the reading passages should be more 'real-world' and have more New Zealand specific content: "I just thought that the readings were not what I expected at all and also that they were not culturally appropriate to New Zealand" (Primary school teacher, F1). For the New Zealand workplace, it seems that reading material based on health and safety practices, work policies and procedures would be most appropriate. As regards writing skills, most native speaker participants (test takers and employers) expected that a language test for New Zealand would contain simple tasks where candidates would be required to write short passages, relating personal feelings, experiences and social or workplace aspirations. The non-native speaker test takers were used to practising IELTS test content. Nonetheless, some of them identified the need for the test content to be more aligned with living and working in New Zealand.

IELTS is an international test, but the consequences of using the same test in a wide variety of contexts should be scrutinised for ethical reasons. After viewing or taking a sample test, the participants in my study were generally sceptical about the relevance of the content of the IELTS test for the New Zealand context. Polio and Williams (2009) suggested that a test that is ethical in one context may not be ethical in another. For these reasons, Uysal (2010) advocated research into whether the characteristics of IELTS tasks match tasks in the target language use domain (TLU) while Lumley and Brown (2005) called for researchers to analyse the test consequences in the context where they are used. Likewise, Knoch and Macqueen (2016) argued that language test developers need to ensure that tests are authentic, context-sensitive and profession-relevant, even when they are used in a variety of contexts. The findings in my research show that the test content of IELTS is not necessarily authentic or profession-relevant for the New Zealand context and workplace.

In general, participants in this study appreciated that a language test would assess the four skills, although listening and speaking were considered to be more important than reading and writing by every participant. Communication with a person rather than a computer was appreciated. However, participants expected simpler reading and writing tasks with New Zealand content.

7.1.3 Test takers and test users are aware of/appreciate the constructs being tested in language assessment in general

All native speakers in this study were overwhelmingly of the opinion that the test was far harder, more complex and more time consuming than they had anticipated.

7.1.3.1 Test takers and test users are aware of/appreciate the constructs being tested in the listening and speaking sections of the IELTS test

Overall, the native speakers (both test takers and test users) in this study appreciated the testing of listening and speaking skills. However, they expected a language test to be assessing listening and speaking in a communicative setting. The constructs they appreciated as being necessary in language testing can be referred to as basic language cognition (BLC), as defined by Hulstijn (2011), since these were regarded as being essential for the workplace. It could be assumed that native speakers of English would have no problems with the speaking section of IELTS as the test developers claim "IELTS assesses the language ability of people who need to study or work where English is the language of communication" (ielts, 2019). As Aryadoust (2013) argued, this statement implies that test tasks are similar to real-life situations. The comments made by the native-speakers of English in this study who work and study in New Zealand demonstrate that tasks incorporating dialogues (the first and third parts of the speaking test) as opposed to a monologue (the two-minute speech in the second part of the speaking module) have a stronger resemblance to real-life communication. In fact, many native speaker test takers found it difficult to speak for two minutes on a topic. This may indicate that performance on a monologue should not necessarily be extrapolated to mean competent performance in non-test situations.

Similarly, the sections of the listening test which were not recognised as being as significant as other sections by the native speaker test takers were listening to a description of a map and a lecture (a monologue). The ability (construct) to follow the description of a map and listening to a presentation or lecture without being able to ask questions was not appreciated by the native speaker test takers because they felt this does not relate to authentic workplace interactions. The crucial point highlighted by the participants in this research is that listening is not a skill which is usually conducted in isolation in workplace situations. This means that score results from the listening section do not necessarily indicate that a person is skilled at listening in workplace situations where a person can use confirmation, clarification and comprehension checks (Loewen, 2015) to elucidate meaning. In addition, the construct of listening for specific lexis in a literal sense is important but not the only skill needed for living and working in New Zealand since inferring meaning from context can

often be an essential skill in authentic language use. Participants in this study also complained that in the listening section they had to remember the exact words stated in the recording and that the marking guide did not accept other correct responses. This is corroborated by Aryadoust (2013), who argued that the listening module of the IELTS test underrepresents listening constructs, namely because the test only assesses the ability to retrieve explicitly stated information and to make paraphrases, and that the marking guides sometimes discount correct answers.

Nonetheless, the non-native speakers appreciated the skills and constructs being tested in the listening and speaking sections test as they saw these as being essential for interacting in an English-speaking environment. This was especially the case after gaining employment in New Zealand. The construct that was remarked upon as being difficult for some non-native test takers was the ability to have native-speaker like pronunciation. Researchers such as Hulstijn (2015), Loewen (2015) and Ortega (2013) have explained the links between pronunciation and identity, and the inability for many language learners to acquire native-like pronunciation post-puberty. If native speakers cannot erase their regional accent since pronunciation is partly physical and speech sounds in a speaker's first language are stabilised between the ages of five and seven (Ortega, 2013), I contend that it is not ethical to assess communication against native speaker pronunciation. Kunnan (2018) and McNamara and Roever (2006) described the unfairness of using pronunciation for boundary maintenance and claim that this has been an artificial method of excluding potential migrants for centuries.

These data, therefore, show that a high standard of oral skills is meaningful for test-takers and test users in the New Zealand workplace. There is strong backing from these research participants that a good performance in the oral sections of IELTS, especially in the sections which are dialogues as opposed to monologues, may have a connection with real life performance. However, native-like pronunciation should not be regarded as necessary or likely to be attained. Ultimately, the concept of native-like pronunciation is not necessarily the same for people from, for example, Scotland and New Zealand.

7.1.3.2 Test takers and test users are aware of/appreciate the constructs being tested in the reading and writing sections of the IELTS test

Language tests are assessing proficiency levels which range from non-existent or basic language skills to highly proficient. These levels need to comprise the abilities of the so-called 'expert' native speaker, who McNamara and Roever (2006) described as being non-existent because native speakers vary widely in their language abilities. For the receptive skills (reading and writing), the

constructs being tested, especially at the higher levels of the proficiency scale, include the ability to comprehend and spontaneously produce texts which are lexically and morphosyntactically complex in a short amount of time. Pilcher and Richards (2017) argued that IELTS assesses de-contextualised, neutral, accurate language which was not valued by the British university lecturers in their study. The native speakers in my study noted that reading texts used widely in New Zealand should be comprehended by a 13 to 15-year old. They did not appreciate the need to read and comprehend texts which contain complex language, and which cover a range of unfamiliar topics. As for written production, they anticipated test candidates would need to write short, simple, easily comprehensible pieces of information.

The comments made by the native speaker test takers, particularly regarding the constructs being examined in the reading module can be summarised by looking at the literal versus interpretative dichotomy, followed by local versus global levels of engagement as analysed in the study by Moore et al. (2011). While the university lecturer (M1) felt that True/False/Not Given questions were more holistic and critical, he found most of the questions to be too literal. This finding is endorsed by Aryadoust (2013) who mentioned the narrow constructs being tested by IELTS. The majority of my native speaker research participants remarked that the question types entailed readers looking for specific detail and both the tasks and topics were not related to the target language domain, which in this case is the New Zealand workplace.

The writing section attracted exactly the same remarks from the native and non-native test takers, namely that they had a lack of knowledge of the topics. Time pressure and the fact that the writing section comes at the end of the day were also issues which affected performance in this section. The non-native speakers' lack of knowledge about the topics could be attributed to diverse socio-cultural educational norms. This aligns with the difficulties faced by many Asian students I have taught who have not understood the topics for Task Two writing in the IELTS test, even though they seem perfectly clear to me as a teacher. This may not seem to be an issue for native speakers of English and yet the topics, together with a lack of time for proof reading did prove to be problematic for native speakers in this study. The native speaker test takers and the employers commented that technology is commonly used to aid writing in the modern workplace. Knoch and Macqueen (2016) suggested that technology is changing how language is used by professionals and that testing systems need to keep up to date with such practices to be valid. The viewpoints of employers are under-researched. However, the employers in this study did not see the relevance of the writing topics or high standards of written work with syntactical and lexical depth. This is also reflected in studies by Gribble et al., (2016), Moore et al. (2015), and Pilcher and Richards (2017).

Overall, the reading section was seen as being important, but as overly complex and lacking Australasian cultural significance by some of the research participants. This implies that there was partial support for the reading section of IELTS. The permitted answers in the reading and listening sections were sometimes deemed to be too literal and constrictive. The writing section proved to be difficult for the native speaker test takers and was regarded by test takers and employers as being inappropriate for the workplace. The constructs being assessed in the writing section had little meaning for native speaker test takers and employers in my research. Affective responses showed that the whole test was much harder than native speakers of English had expected.

7.2 RQ2: How well do test takers and test users feel they can interpret and use test score results (decision inference)?

A test can be considered to be valid if test takers and test users can use and interpret test score results. IELTS claims that organisations around the world rely on IELTS to help them select the right people in a wide range of sectors such as finance, government, construction, energy, aviation and tourism (ielts, 2019). The research participants in this study worked in similar employment sectors, so their ability to interpret and use test score results is pertinent.

7.2.1 Test takers and test users feel they can interpret and use test score results

The test score results had little meaning for the native speakers of English who were test takers especially the results for the reading and writing sections. The test taker (M4) involved with teaching IELTS had the most knowledge about the mechanics of the exam and yet, even for him, how the writing section is assessed was “as clear as mud”. The majority of the native speaker test-takers were surprised by their results in the reading and listening sections; they expected they would score highly since English was their first language. The native speaker test takers perceived the writing samples composed by non-native speakers presented to them as being acceptable for communicative purposes in the workplace. While non-native speakers of English reported learning test taking strategies to improve their test score results, they also realised that these strategies did not translate into competence to live and work in an English-speaking environment. The two non-native university students (F GNNs and M SNNS), the PhD candidate (F MNNS) and scientist (F FNNS) found no connection between their achievements using academic English and their IELTS writing scores.

7.2.2 Test takers and test users (New Zealand employers) feel they can interpret and use test score results

Employers could see no correlation between test score results and workplace competence. This perspective was echoed by the university employer (Emp 1) who had noticed that the writing score results for prospective non-native teachers were consistently lower than their results for the other three modules, but he could not understand this trend. It is important to note that most of the employers from the education sector knew which band scores were required in their area, but they had very little awareness of what these band scores represented in terms of language competence. It is pertinent to note that the senior educator (Emp 2) was under the impression that an IELTS 7.0 indicated just basic competence. Yet after viewing a sample test her perception changed dramatically; her new perception was that this type of test would be difficult for native speaker teachers. Employers in the IT, business, manufacturing and retail sectors were, in general, mystified by the band score requirements for immigration or professional registration (6.5 for immigration and 7.0 for professional registration), and the written samples they were shown were regarded as being either adequate or more than adequate for their workplaces, even though these samples had scored between 5.5 and 6.5. Interestingly, the sample written by the non-native speaker who scored 6.0 was favoured by 80% of the employers in this research.

Ultimately, in this study, none of the stakeholders appeared to understand the scores and what they could infer about a prospective employee's linguistic abilities. As Spolsky (2008) pointed out, lay people often incorrectly think that language testing measurement is simple. Unidimensional scores can look valid and reliable, but they can belie the competence and hence the fate of test takers. Brewer, Knoeppel and Lindle (2015) stated that the public falsely expects testing to be precise, but intelligence in general, and language competence in particular, is not one-dimensional with the result that perfectly capable people may not be selected for academic or work purposes. The ESOL teacher (M4) who taught IELTS preparation classes was mystified by the scoring for the writing module. He had taught students who were aspiring to become nurses in New Zealand and felt that their writing was competent, and yet, they could not gain higher than 6.5 in the writing section. The university lecturer (M1) had dealt with international students and could not perceive any correlation between their IELTS scores and their linguistic competence in his courses. The business employers who were regularly presented with IELTS scores when migrants applied for jobs, could also see no relationship between the scores and an applicant's linguistic competence at work. There was a stronger relationship between the listening and speaking scores than the reading and writing test scores and workplace competence. This could be why employers in the research by Gribble et al.

(2016) perceived employees in the healthcare and early childhood sectors as being competent, and yet these employees struggled to gain the required IELTS test scores. In short, this study has supported the claim made by Elder and Kim (2013), who argued that setting cut-off scores to maintain professional standards may result in inaccurate and unethical decisions.

Who should ensure test scores can be used and interpreted appropriately? Johnson and Riazi (2017) state that "test users in particular are often if not typically negligent in investigating both the meaning of test outcomes and the impact of test use for their particular educational context" (p.85), whereas other researchers (Kane, 2013) claim that it is incumbent on test developers to ensure tests are used ethically. As in my research, Pill and Harding (2013) discovered that gatekeepers, such as government officials and medical representatives, had very little language assessment literacy even though they were making decisions which could have life changing outcomes based on test score results. Ultimately, I contend that the onus should not lie on neither test developers nor test users; there needs to be more engagement between test developers, policy makers and language testing researchers to ensure testing systems are being used in a socially and politically appropriate way (Chalhoub-Deville, 2016; Pill & Harding 2013). I would also argue that the economic consequences for test takers and employers are also essential for validity arguments, although economic impacts have not often been investigated.

In brief, the test score results had little meaning for the native speaker test takers and test users, since they could not interpret the test scores. The non-native speaker test takers had mostly had prolonged experience with IELTS and therefore seemed to understand the test score results. The one section they could not interpret was the writing score results as these scores bore no relation to their written competence in their places of work or study. The findings indicate that the score results for the oral sections may be useful but the score results for the reading section and, in particular, the writing section did not appear to be useful for test-takers or employers. These findings, therefore, demonstrate that language assessment is a complex area as highlighted by many other researchers (Aryadoust, 2013, Chapelle et al., 2010, Loewen, 2015, Ortega, 2013, Pill & Harding, 2013, Ricento, 2014, Spolsky, 2008), and the claim that test score results can be validly used by stakeholders to indicate workplace competence may not be warranted.

7.2.3 What do test users feel about the required minimum standards for L1 and EAL test takers?

The assumption that test cut-off scores are appropriate for New Zealand was analysed by looking at test scores of native speakers of English. The fact that native speaker test takers did not score highly

in a sample test highlights the unfairness of expecting non-native speaker migrants to gain higher scores in a language test than the average English speaker. Comparing the non-native speakers with native speaker test results in this study, there were not great discrepancies in the speaking and listening modules; many of the non-native speakers had received good results for the speaking and listening sections. The native speakers fared reasonably well in the listening module with scores between 6.5 and 8.5. However, the native speakers' scores for the reading section were lower (between 6.0 and 8.0). The non-native speakers in this study felt that the module where their scores were not compatible with their abilities was the writing section. Their writing score results can be compared with the native speaker test takers in my research who scored between 4.0 and 8.0. The average score of the native speakers for Task One in the writing module was 5.58 and for Task Two 6.29. Many native speakers of English scored below 6.5 (a score needed by non-native speakers for immigration) for their writing skills and well below 7.0 (a score needed by non-native speakers for professional registration). In fact, only one native speaker in this study scored 7.0 or above for Task One writing while five scored 7.0 or above for Task Two writing. The test results for the native speaker test takers included fifteen scores (across both tasks) between 4.0 and 6.0 for writing. None of the native speaker test takers scored the perfect 9.0. Since non-native speakers of English are required to attain 6.5 or 7.0 for immigration and professional registration purposes respectively, I contend that present cut-off scores are not being used in an ethical way.

Native speakers differ in cognitive abilities, literacy, levels of education and in their professional backgrounds (Hulstijn, 2011; Stricker 2004). From a usage-based perspective of language acquisition, native speakers do not all have the same extensive input or output in the reading and writing modes of language. This could be the reason why the IELTS scores of the native speakers in this study were lower in the reading and writing modules. This trend may be exacerbated by changes in the workplace where written output is no longer expected to be extensive, complex and accurate at all times.

Participants in this study welcomed the objectivity of using outside institutions with trained raters who focus on linguistic proficiency in a cognitive way. In this respect, IELTS is a testing system which is reliable and defensible (Ryan, 2016) although I would contend that it is still not transparent, especially in its dissemination of the meaning of test scores. This is discussed further in 7.7. Some researchers (Knoch & Macqueen, 2016) have explained that when tests have exempted native speakers on the assumption that they will gain the highest scores, such tests are not being used in a fair or equitable way. Further, Knoch and Macqueen (2016) and Pilcher and Richards (2017) argued that a combination of language trained raters and domain experts may make the best judgements

regarding competence for the workplace. I would suggest that this may be complicated in practice. However, I agree with Hulstijn (2011) that more transparency about what a language score at the higher levels means, and more research into the range of scores most native speakers are likely to attain may help alleviate any unfairness and the unethical uses of test score results.

In response to RQ2, the non-native speakers of English viewed scores for the speaking and listening modules as being beneficial for living in New Zealand whereas the results for the reading module were viewed as being partly beneficial. Not all native speaker test takers scored highly in the reading module, but their scores were noticeably lower in the writing section. This may be because there is too much emphasis on the traditional rules of form in the IELTS marking rubrics, rather than meaning and intelligibility, as claimed by Uysal (2010). For this reason, the decision inference represented in RQ2 cannot be supported by the data in this study. Developers of high-stakes language tests may need to revisit the marking rubrics for the writing module to align them with current writing practice in the workplace and provide test users with more language assessment literacy regarding the likelihood of native speakers gaining (or not) high scores in the reading and writing sections.

7.3 RQ3: How well do test takers and test users feel that the IELTS tasks and the constructs being tested reflect communicative language used in the New Zealand workplace (extrapolation inference)?

IELTS claims that IELTS test scores are a reliable gauge of test takers' ability to communicate in English. To ascertain if IELTS assessment tasks reflect communicative language used in the New Zealand workplace, data from interviews with test takers and test users were used to see if the research participants believed that IELTS assessment tasks and the constructs being tested reflect the type of communicative language used in today's workplace. In other words, do test-takers' and test-users' comments support or rebut the fact that test tasks and test performance correlate with linguistic competence in the NZ workplace. This extrapolation inference is particularly relevant when tests are used for purposes they were not originally designed for (Knoch & Macqueen, 2016). Research has been conducted for these purposes into the Occupational English Test (OET) by Elder, Harding and Knoch (2009) and Macqueen, Yahalom, Kim and Knoch (2012). However, these studies focus on the healthcare sector, and not on the changing role of IELTS, which is now promoted just as much for business and immigration purposes as for academic entry to tertiary institutions. As discussed by Kane (2013) and Knoch and Macqueen (2016), it is imperative for validation research to be ongoing.

7.3.1 The characteristics of the oral assessment tasks correspond to tasks required in the workplace

As stated in Chapter 4, my research participants expected a language test to be focusing on basic communication skills. This may be because oral communication is valued above every other skill in the workplace. Indeed, McNamara and Roever (2006) explained that in any speaking test, language is co-constructed by interlocuters. In line with skills identified as being essential for the 21st century workplace (critical thinking, creativity, collaboration and communication), participants in this study commented mostly on the need for effective communication, followed by the importance of collaboration and critical thinking as opposed to following orders. In addition, all the employers I interviewed explained that cultural norms such as the propensity for employees to question instructions were more important than language skills. Creativity was not remarked upon so much, although this could be deemed as being a part of problem solving, which was mentioned as being essential by all the employers in this study. Tasks required in the workplace include customer service, and dealing with colleagues as well as people outside the organisations, such as parents. Oral communication at work is not generally formal, but rather semi-formal, with pragmatic competence for New Zealand society being rated as important by New Zealand employers. This includes knowledge of New Zealand idioms, slang and an understanding of the New Zealand accent. It is important to note that according to some of the employers, cultural issues specific to the New Zealand context play a role not only for non-native speakers but also native speakers of English from different English-speaking countries such as the USA.

7.3.2 The characteristics of the reading and writing assessment tasks correspond to tasks identified by employers as being important for the workplace

Reading and writing may be counted as important skills in all workplaces, but it appeared from the participants' comments that reading and writing tasks have changed in the last couple of decades. The result of this was that the IELTS reading tasks were not generally seen as being pertinent, even in the educational sphere. This is because the topics were seen as being too scientifically and technically oriented, and the amount to be read in a short time was deemed to be excessive. Native speaker test takers and test users felt the need in the test to read the questions and flip back and forth between questions and text was an artificial skill. The reading tasks in today's workplace are generally emails, processes and procedures, handbooks and reports. This finding is supported by the research undertaken by Moore et al. (2015) in the Australian context. All participants mentioned that it was crucial for employees to comprehend health and safety rules. As for writing, the ability to produce an essay was not a task that was identified by any of the employers as reflecting

employability. It was noted that describing graphs may be useful in a business sphere, although this was not valued by the majority of the participants in this study.

7.3.3 Constructs being tested are the language constructs needed for the TLU domain

All participants involved in this research viewed a language test, such as IELTS, as assessing cognitive academic language proficiency as opposed to basic interpersonal communication skills in all four skills (speaking, listening, reading and writing), but especially in the writing section. One of the reflections that I did not expect as a language teacher and examiner was that a test of this kind would be assessing educational level at an unattainable level for native as well as non-native speakers of English. This was because I had always assumed language tests were assessing general language competence. Surprisingly, the non-native speaker test-takers in this study seemed to be more aware than teachers in general that educational level can have a strong effect on language proficiency.

Many non-native speaker test takers remarked that this type of test would be difficult for many inhabitants in their home countries if they were taking the test in their first language. Indeed, Hulstijn (2015) contended that many adult native speakers of any language would not be able to attain the C1 (IELTS 7.0 - 8.0) or C2 (IELTS 9.0) levels on the CEFR framework (Common European Framework). For a comparison of the CEFR and IELTS band score results see Appendix R. Likewise, Ricento (2014) found that an idealised, perfect form of language used by a homogenous native speaker does not exist in most areas of society. Hulstijn (2015) suggested that research needs to be undertaken to ascertain at which CEFR levels the linguistic abilities of native speakers start to differ to gain a better understanding of which levels represent basic language cognition (BLC) and where higher language cognition (HLC) starts to be visible. Although Hulstijn (2015) could only surmise that the language abilities of native speakers may start to diverge around the A2 (IELTS 4.0) or B1 (IELTS 4.0 – 5.0) levels, my research shows a demarcation starting to emerge around the IELTS 5.0 – 6.0 (CEFR lower B2) level, especially in the writing section. When native speaker test takers and test users (employers) viewed four sample writing responses with scores of 6.5 (NS), 5.5 (NS), 6.0 (NNS), and 5.5 (NNS), all of the samples were deemed to be similar or of a higher standard than writing that is commonly produced in New Zealand workplaces. The writing sample that employers in this study felt was most representative of the writing in the workplace was the Task Two piece written by the accounts manager (F5) who received a score of 5.5. This indicates that a score lower than IELTS 6.0 may represent basic written communicative competence whereas a score above IELTS 6.0 indicates a test taker has written language competence acquired only through extensive involvement in

academic or professional contexts. A further point to consider is that the native speakers in this study were not blue-collar workers; they were employed in the education, business, IT and customer service areas and many considered their English to be of a good standard. This implies that the rubrics for the writing section, which are used by trained IELTS raters for marking purposes, may not reflect authentic language as written by native speakers in today's workplace. As such, test score results for non-native speakers of English may not be being used ethically.

7.3.4 Domain experts value the constructs being tested

It appears from my study that in today's workplace precise, sophisticated language is no longer expected. Even managers in businesses that liaise with government departments explained that complex language used to be expected, but that has changed in the last decade with simple, clear language now being valued. This finding was corroborated by the research undertaken into workplace English by Moore et al. (2015). The ability to summarise, and to convey a message in as few words as possible is now paramount (Moore et al., 2015). Many participants in my research explained that when language was complex, this could cause problems since the readers may not understand the message. In today's workplace, employees receive large amounts of information and need to process it quickly. Therefore, the need for basic language cognition outweighs the need for higher language proficiency. This is reflected in the research undertaken by Hulstijn (2011). Likewise, according to Pilcher and Richards (2017), the university lecturers in their study did not expect perfectly accurate language; rather they wanted language which was comprehensible. For Pilcher and Richards (2017), the constructs being assessed by IELTS are mistakenly assumed to help with academic success. My research has indicated that the ability to gain high scores in writing section of the IELTS test can mistakenly be assumed by test takers to guarantee employability and workplace success.

One clear finding emerged from analysis of the interviews in my study; the writing section was not related to either workplace or academic literacy skills because it seems that a) how language is used for written correspondence has changed, and b) high levels of language proficiency with grammatical accuracy, correct spelling and punctuation are no longer expected in the modern workplace. This points to construct-irrelevance (McNamara & Roever, 2006) when tests are not typical of real-world discourse. The data indicate that test results may be meaningful in terms of cognitive academic language proficiency, but that test performance, particularly in the writing section, may not be extrapolated to mean language competence for modern workplaces or even success as a student at tertiary level.

As regards RQ3 it appears that oral communication as tested by IELTS may be necessary in the New Zealand workplace. However, language use in isolation is eclipsed by the need for collaboration, critical thinking and sometimes creativity. Accuracy is no longer as important and, what may not be appreciated by decision makers, is that native speakers may not attain high levels of higher language cognition (HLC) as defined by Hulstijn (2011; 2015). My data have demonstrated that the extrapolation inference can be supported for listening and speaking modules, partially supported for reading module, but not supported for the writing module.

7.4 RQ4: What do test takers and test users perceive to be the consequences of test use for the New Zealand workplace (consequences inference)?

IELTS claims that the test is fair to all, and that it is recognised and trusted worldwide for its fairness, reliability and high-quality standards. For this reason, the perceptions of test takers and test users in New Zealand were used to see if the test is perceived to be fair and beneficial for stakeholders in New Zealand. Test consequences are beneficial to all stakeholders (test takers, test users and test developers) if they promote test takers' and test users' best interests. According to Kane (2013) the evaluation of consequences, both intended and unintended, plays a crucial role in validating score uses. Similarly, Kunnan (2018) argued that assessments "ought to be evaluated in terms of whether they can bring about beneficial consequences to test takers and the communities in which they are used" (p.200). Research into test consequences is particularly essential because Brewer et al. (2015) claimed that they are often neglected.

7.4.1 Decisions made regarding the test and score results are beneficial for non-native speaker test takers?

The non-native speakers in this study mostly appreciated the use of tests such as IELTS for migration to New Zealand. They valued the reputation of IELTS and were, in general, appreciative that they had received the required results to work in New Zealand. The test was mostly beneficial for them, especially the speaking and listening sections. They were of the opinion that attaining a good score in these oral skills helped them live, study and work in New Zealand. The manufacturing manager from India gained confidence from his IELTS test score results. While some non-native test takers appreciated the reading section, most were discouraged by their results for the writing section and hence lost confidence in their language abilities. This led some to avoid writing as much as possible and feeling as if they could not realise their dreams, even if they had tertiary level qualifications. Additionally, some test takers such as the primary school teacher from Germany, the librarian from Poland and the manufacturing worker from the Philippines were unable to gain professional

registration or work because of their test score results in the writing section, despite working and having received tertiary qualifications in New Zealand. In all three cases, their listening and speaking skills were higher than the results from the writing part of the test. Because a score of 7.0 is needed in all four skills, the writing test score results hindered their employment prospects. The teacher from Germany was fortunate to have her case reviewed by the Education Council Aotearoa, but the librarian from Poland could not gain entry to a teaching course in New Zealand despite having a Master's degree in Physical Education from her home country. This research participant (F PNNS) had taken the IELTS test approximately ten times and felt the ongoing costs (mental and monetary) were not worth her while. This situation was similar for the manufacturing worker from the Philippines who had worked in New Zealand for three years and was finding the costs of taking IELTS prohibitive. These findings are echoed in the research by a number of authors (Chalhoub-Deville, 2016; Kunnan, 2018; Pilcher & Richards, 2017; Spolsky, 2008) who bemoan the scale and lucrative nature of high-stakes language tests for test takers, especially when they are used for new purposes, such as business and migration purposes as opposed to university entrance. Indeed, participants in the educational sector (F1, M1) recognised that native speaker test takers would probably have to pay for coaching to be able to gain an IELTS 7.0 and that this would likely result in a lucrative training industry if they needed to take this type of assessment.

It cannot be ignored that financial costs were only one consequence for non-native test takers. The mental anguish of failing to gain employment after taking the long and arduous journey to migrate to a new country is a serious social consequence, especially when employment is denied for a point or even half a point in one section of a language test (writing) which native speakers may struggle with.

7.4.2 Decisions made regarding the test and score results are beneficial for non-native speaker test takers?

One aspect of high-stakes language testing that has been under-researched is the fact that candidates who have English as their L1 may, in certain circumstances, need to take a test such as IELTS for migration or professional registration purposes. The native speakers who took a sample IELTS test in this research all felt they were competent users of their L1. Although the native speaker test takers felt that the IELTS test was rigorous and reliable, they felt that the difficulty of the reading and writing sections made the test unfair for prospective migrants to New Zealand. Further, they perceived the language in the test as being unfair for them because it was unrelated to either New Zealand life or the New Zealand workplace. A few of the native speakers in this research felt that the answers in the reading and listening sections were so prescriptive that their own responses were,

therefore, unfairly categorised as incorrect. Other aspects of the sample test they took that they felt was unfair was the need to speak for two minutes uninterrupted in the speaking section and to listen to lectures while completing an answer sheet since these were competencies they did not need in the workplace. Overall, the test was deemed as being unfair for L1 speakers of English who work in sectors such as construction, IT, manufacturing, retail and tourism because many native speakers who work in such areas do not need to use complex language with accurate spelling and punctuation. A few of the L1 test takers were disgruntled by their reading and listening test score results because they expected to gain high scores, especially because English is their mother tongue.

7.4.3 The use of language testing is beneficial for employers

The New Zealand employers did not find language testing results to be beneficial for them. This was partly because the test score results had little meaning for them and bore no correlation with the skills of migrants they employed. While the employers in the education and business/IT sectors felt that they were losing valuable employees because they could not attain the required test scores, the situation for the retail and manufacturing sectors was more serious. When they could not employ New Zealand staff, employers in these sectors stated that they needed to seek overseas employees. Yet, immigration policies which require migrants to gain 6.5 IELTS in all four skills meant that they found it difficult to hire and retain staff members from overseas. This led some of these employers to comment that language testing was proving to be "devastating" for their industry. This implies that business continuity and productivity in New Zealand may have been affected by using language tests for immigration purposes. However, it must be acknowledged that language testing is only one of the measures imposed by immigration officials in line with immigration policies. Other criteria include work experience and the relevance of a migrant's work skills for the New Zealand workplace.

7.4.4 The use of language testing is beneficial for New Zealand society

Language tests are often used as gatekeeping devices (Knoch & Macqueen, 2016; Macqueen, Pill & Knoch, 2016). However, participants in my study questioned what type of people would be allowed to migrate New Zealand. Some participants pointed out that every person has different strengths and weaknesses with some being linguistically adept in the workplace whereas others may be linguistically challenged. Nonetheless, this does not necessarily relate to workplace competence. As many participants pointed out, managers and directors of companies do not necessarily produce complex or accurate English in today's workplace. Many participants also noted that New Zealand does not need only high functioning migrants who are linguistically competent since that would not generate a 'normal' society where it takes all sorts of people with many different skills to produce a

productive whole. Participants argued that New Zealand needs builders, artists and software developers, who need technical, artistic or coding skills, not language skills (this applies to native as well as non-native speakers). The danger is that valuable workers could be excluded due to poor writing skills that are not actually needed in their respective workplaces.

The fact that employees, or indeed employers, may have dyslexia was mentioned but the research participants generally described these staff members as being often more than capable. This view was highlighted in the study by Pilcher and Richards (2017) where a lecturer in Product Design felt that if IELTS assessed accurate language, it was not beneficial for students on his course. This was because some of the best students on his course may be dyslexic and therefore fare badly in a test like IELTS, and yet they were often exceptional designers.

The primary school teacher (F1) in my study observed that any policy that used this type of test could lead to a racial bias or, in her words, "the white is right" phenomenon. This aligns with research by Kunnan (2018) and Ricento (2014) who argued that policies are often political in nature, and although the world has become a global marketplace (Spolsky, 2008), language tests are often unethically used as boundary objects (Kunnan, 2018; Macqueen, Pill & Knoch, 2016) to keep out 'undesirable' nationalities. Conversely, many of the native speaker participants valued the skills of multilingual migrants for New Zealand. This builds on research by Ricento (2014) who described the danger of governments discriminating against people who do not speak English. It is often the case that such people are seen as being deficient rather than proficient because they can speak more than one language.

Most participants in my research study noted that, for them, it was clear that policy makers have very little awareness of language testing and what can be inferred from the results. This finding corroborates those by Pill and Harding (2013) and Ricento (2014) who argued that political theorists need to understand that languages are not stable, nominal constructs and that most people do not use an idealised variety which is based on the written language of the most educated native speakers. Indeed, error-free English in all contexts is "unrealistic, irrational, and counter-productive in a world in which the users of English in the outer and expanding circles far outnumber those speakers on the Inner Circle varieties" (Ricento, 2014, p. 364). In my research I have endeavoured to respond to the call by Chalhoub-Deville (2016) who argued for socially centred validity explorations and, thereafter, engagement in policy research.

Overall, IELTS test consequences are mostly beneficial for non-native speaker test takers. However, the test results in the writing section proved to have negative consequences which can regarded as

being unethical. One unintended consequence was that the non-native speakers were often barred from employment and/or professional registration for the sake of half a point, as in the case of the teacher from Germany who could not register with the Teaching Council of Aotearoa New Zealand because she achieved only 6.5 in the writing module. Paradoxically, she attained an A+ at a New Zealand university when studying childhood literacy. According to the New Zealand employers in my research, a second unintended consequence was that the use of test scores by immigration authorities and professional bodies led to the exclusion of competent employees in the education, manufacturing and retail industries, which corroborates the research by Gribble et al. (2016). This has been 'devastating' for some industries according to company directors I interviewed. The consequences of using language testing may not be positive for New Zealand society in general since a variety of employees are needed in every society, not just high functioning, linguistically competent people. A few employers and test takers in this study noted that using test score results could lead to intended or unintended racial biases. These viewpoints about language and language policies being used to exclude migrants in a racial way, and which can lead to the underemployment of migrants relative to their pre-migration skills, support the findings of racial biases towards migrants in Australia reported by Butorac (2014), Kunnan (2018), Major, Terraschke, Major & Setijadi (2014), Roshid and Chowdhury (2013), and Scarino (2014).

7.5 Overview of responses to the research questions

The validation argument approach used to analyse the perceptions of native speaker and non-native speaker test takers, and native speaker employers of migrants, was presented in Chapter 2. The results, as discussed in 7.1 to 7.5 show that language tests are generally meaningful for my research participants and relevant for New Zealand. Test score results from language tests are mostly being used validly in the listening and speaking modules, but performance in the reading and particularly in the writing cannot be extrapolated to mean linguistic competence in the workplace. Native speaker test takers and native speaker employers could not interpret or use test score results, which means that decisions about a migrant's employability based on language test results in the reading and writing modules are not being made ethically. The economic and societal consequences of language testing for New Zealand employers, as reported by the test users in this study, were largely negative.

7.6 Implications

Chalhoub-Deville and Deville (2005) called for investigations into how test scores are reported. I contend that it is more crucial to understand the kaleidoscope of inferences made about the test itself by a variety of stakeholders, the interpretation of scores and the consequences for test takers and test users since these have impacts for society which can be far-reaching. Lumley and Brown (2005) stated that tests are not isolated events; they are social and political tools whose impact must be evaluated as part of a validation process. The economic implications of using high-stakes language tests have often been neglected. For these reasons, the social, economic and political implications will be discussed in this section.

Looking at the social implications, non-native speakers of English who migrate to an English-speaking country are sometimes deemed to be deficient by the predominantly monolingual English-speaking majority because they do not speak English 'perfectly' (Butorac, 2014; Major et al., 2014; Roshid and Chowdhury, 2013; Scarino, 2014). The IELTS rubrics used for English language assessment exacerbate this situation because high scores are attributed to test takers who use a wide range of lexical and grammatical structures in a sophisticated and natural manner with only rare minor errors or slips (appendices O and P). This means that to attain 6.5 or above for migration and professional registration purposes, it is necessary to have higher language cognition (Hulstijn, 2011). Many of the native speaker test takers in this study had strong educational and professional backgrounds, and yet they did not necessarily score highly in the IELTS test, especially in the writing section. As mentioned above, the wording of the rubrics used by IELTS raters to assess the written scripts means that complete accuracy and complex language are valued (Uysal, 2010). Yet, such language is not expected, or commonly produced, in today's workplace. Ricento (2014) discussed the danger of governments discriminating against people who do not speak English, but the situation is worse if non-native speakers need to gain much higher scores that the average native English speaker is likely to attain. For this reason, New Zealand may be excluding competent migrants resulting in an unbalanced mix of migrant workers, a finding supported by Brewer et al. (2015), Cizek et al. (2008) and Davies and Elder (2005). From a societal point of view, New Zealanders also need to unlearn monolingualism (Scarino, 2014), and make the effort to understand people from other countries who have spent many years learning a second (or third) language for the country to prosper (financially and socially). As Spolsky (2008) has pointed out, fitting people into boxes in an arbitrary Procrustean bed manner may not benefit society as a whole.

Economically, the employers in this study noted the need to employ migrants and the costs for businesses when they spent time and effort to train staff who were then not allowed to remain in

the country. The employers were perplexed because they needed to employ and train migrants, but if prospective migrants could not achieve the required cut-off score (6.5) for a work visa, these migrant employees had to leave New Zealand, resulting in a loss of income and low productivity rates for their businesses. The employers in this research frequently mentioned that they had to keep training new staff for the above reason. The non-native speakers I interviewed mentioned the financial costs they have to incur when they have to study for and take a language test numerous times, and then fail by a narrow margin, even when they have gained tertiary qualifications in English. Kunnan (2018) asserted that fair assessment practices allow test takers to have basic freedoms and the right to be treated with respect. As Pilcher and Richards (2017) argued, it is normal for professional bodies to set cut-off scores to ensure professional standards are kept. Indeed, this was appreciated by many participants in this study. Nonetheless, it is potentially dangerous when professional bodies falsely assume, without knowledge, that the English being tested in IELTS is the English needed for work or study (Pilcher & Richards, 2017; Pill & Harding, 2013). The educator in this study who teaches IELTS to students wishing to work or study in New Zealand was not aware that language tests are assessing higher language cognition (Hulstijn), which is worrying. If language teachers and trained raters are not aware of the scope of language testing and what can be inferred about a person's competence in English based on test score results, it is even less likely for domain professionals (employers and policy makers), who are not expected to have in-depth knowledge of language competence (Pill & Harding, 2013), to have such awareness.

Politicians have used language tests to bolster their political agendas throughout history (Kunnan, 2018; Ricento, 2014; McNamara & Roever, 2006). As suggested by Hulstijn (2011; 2015), policy makers should set reasonable standards based on evidence of which scores native speakers who are successfully integrated in the workplace in New Zealand are likely to attain in the reading, and particularly, in the writing module of IELTS. This study has demonstrated that there needs to be greater understanding of what is being tested and more flexibility regarding which scores are acceptable for today's workplace. It may be politically, economically and socially advantageous to invest in a New Zealand or Australasian specific test of communicative competence. It is noteworthy that Moore et al. (2015) researched the types of language needed in professional domains in Australia and recommended an alternative form of IELTS: one for professionals seeking to work in a business environment. It may be timely to introduce English for professional purposes tests, as described by Knoch and Macqueen (2020), for a wider range of employment sectors. Moore et al. (2015) did not look at the viability of an alternative test for the Australian/New Zealand market so this is an area which could benefit from more research. Another possibility is to include domain

insiders (test users) in IELTS standard setting procedures as described by Knoch and Macqueen (2020) and O'Neill et al. (2007) for the aviation and healthcare sectors. This would allow test users to set more reasonable cut-off scores for their particular fields of employment.

To enhance fairness in society, it may be the case that all applicants (native and non-native) should be required to take a language test for professional registration purposes, especially in reading and writing which represent higher language cognition (HLC). This may have positive consequences for New Zealand society since the majority of employers in this study were more concerned about the literacy level of native speakers of English they employed, rather than non-native speakers of English.

It is hoped that my research may ultimately help decision makers and test developers by encouraging engagement and more thorough discussions regarding the meanings, and outcomes of language assessment; in short, to promote language assessment literacy amongst important stakeholders. As Davies and Elder (2005), Lo Bianco (2018) and Pill and Harding (2013) explained, policy officials and language experts do not often agree on how to implement high-stakes language testing into national or regional policies, but there is a need for researchers in language testing to be involved in the dissemination of increased language assessment literacy. Similarly, Chalhoub-Deville (2016) has called for emerging research to engage in theory of action research (TOA) which includes consequences in validity research and uses a reform-informed validity framework to investigate "the interconnections between policy mandates, testing research and societal consequences" (p. 454) because, as she has stated, publications in this area are scant.

7.7 Research limitations

This research study investigated the experiences of native and non-native speaker test takers. However, there are certain limitations associated with the testing of native speakers. The native speaker participants who responded to a call for research into language testing did not take a live test but, rather, a sample IELTS test. This meant that they were not under the same pressure as normal test takers because their employment or immigration status did not depend on the test results. However, all the native speakers were reluctant to put their pens down when the allocated time for the reading and writing sections was reached, which may have indicated that they did take the test seriously. While they took all sections of the test, they were not graded on their speaking performance. This was because IELTS examiners are not allowed to award test scores to prospective test candidates outside the real test. The

research participants were also not given their writing test results because my main focus was on their impressions of the speaking and writing sections, rather than their actual scores. Since native speaker test takers do not necessarily gain the two highest bands (8.0 or 9.0) in speaking, this is an area which could benefit from more research. As discussed in the methodology section (3.7.2), it was decided early in the research to use a sample Academic test, not the General Training test, mainly because the Academic test is used for professional registration while immigration rules allow for either test to be taken. The native speaker results can, therefore, not be taken to necessarily indicate actual performance in the reading and writing sections of the General Training test, only likely performance. More research into native speaker performance on the General Training test would be beneficial. Lastly, the writing scripts were graded by two experienced IELTS examiners, but it is impossible to state that these would be definitive results.

A further limitation is related to my insider status having been an IELTS examiner for 22 years. It was critical that the research questions were open-ended and that they were previewed and deemed acceptable by the Human Ethics Committee in advance to avoid any unintended bias. While the interview questions were open-ended (Appendices I, J and K), participants commented on topics that were important to their experiences and backgrounds, and many of their perceptions were generally not anticipated by me as a researcher who had been involved in language testing for over two decades.

Additionally, this study investigated the perceptions of employees and employers in three employment sectors only (education, IT/business, and manufacturing/retail). It is important to mention that the viewpoints of the test takers and employers in this research from the aforementioned employment sectors may not be fully representative of the professions they represent. These participants also proffered opinions mostly from a lay perspective. The opinions of domain insiders who have some language assessment literacy may be helpful in future research. In this study the only participants with some language assessment literacy were the ESOL teacher (M4), the teacher from Malaysia (F MNNS) who was an IELTS teacher in her own country and the business consultant (Emp 6), who had taught IELTS listening skills at a local university prior to moving into business. While an analysis of perceptions in other sectors may lead to different findings, the changes in the use of written language may be generalised to other areas since this seems to be a societal change encompassing the educational, business and legal sectors as well as less linguistic rich environments such as IT (in some respects), manufacturing and retail. More research into the impact of language testing on test

takers from non-skilled sectors or from specific non-European background cultures in New Zealand may be beneficial for societies in general and for New Zealand in particular.

7.8 Recommendations for using language test results in the future

The purpose of this study was to ascertain if language testing is being used validly and ethically for migration, employment and professional registration purposes in New Zealand. The insights gained from interviewing native speaker as well as non-native speaker test takers, and New Zealand employers have highlighted the following considerations for the future use of high-stakes language tests such as IELTS:

7.8.1 Recommendations regarding the New Zealand domain

There needs to be more transparency from test developers as regards what is being tested in high-stakes language tests, namely if tests are assessing communication or proficiency. All participants in this research appreciated the need for communicative competence in the New Zealand workplace, and for this reason strong speaking and listening skills were valued. However, after viewing a sample test the participants in this research (native and non-native speaker test takers as well as employers) recognised that attaining a high score in this type of test is dependent more on high educational or professional levels rather than the ability to communicate basic personal information in a language.

7.8.2 Recommendations regarding extrapolation to the New Zealand workplace

Test developers and researchers should investigate the extent to which the role of reading and writing in the workplace is changing. Many employees are not expected to read large amounts of complex language, and while writing mostly needs to be brief, clear and simple. Language testing systems are looking for the opposite; language is assessed with the aid of detailed reading texts and exam candidates are required to produce grammatically and lexically complex language which is completely accurate in a short time frame. This implies that the way written language is being assessed in an IELTS exam through the testing rubrics is not convergent with what is valued in today's workplace. Test developers would be advised to reflect on how to test and assess language, especially reading and writing skills, for the contemporary workplace.

7.8.3 Recommendations regarding decisions for workplace competence

Standard setting procedures should involve test users (domain insiders) in determining which level of performance on a test should be set as a minimum requirement for their professions. The native speaker test takers and test users (employers) were unable to interpret test score results in any meaningful way. A senior educator commented that for her a score of 7.0 meant that a person

would be minimally competent in linguistic terms. However, after seeing a test she declared that a 7.0 would be difficult for native speakers to achieve. Written samples which had scored between 5.0 and 6.0 were described as being perfectly acceptable pieces of writing for the workplace. This is despite the fact that a score of 6.5 is needed by migrants to New Zealand and 7.0 for professional registration. None of the employers in this study could see any correlation between test scores and the type of language competence required for the workplace. Rather than rely on language test scores, they preferred to use their own screening processes, and would not accept what they deemed as low oral skills at the interview stage. The information on test scores and how to interpret them (ielts, 2019) had not been accessed by the employers in my research. Test users, therefore, need more information from test developers, delivered in different formats, in order to use test score results effectively.

7.8.4 Recommendations regarding native speaker test candidates

Language testing researchers and policy makers should investigate in more depth the scores that native speaker employees in different employment sectors are likely to attain in this type of high-stakes language test. The majority of native speakers will likely gain scores which are 7.0 or above in the speaking and listening sections, but not a 9.0 without preparation or training. The IELTS website (ielts, 2019) encourages native speakers to prepare for the test and the test statistics for 2018 indicated that only 2% of test takers who have English as their first language scored 9.0 with their mean overall score being 6.92. My research data show that native speakers may score 6.0 - 6.5 or more in the reading module but are likely to score around 5.5 – 6.0 in the writing module. For test scores to be used in valid and ethical ways, authorities should set test score prerequisites for immigration and professional registration at similar levels. Prospective migrants, especially from non-skilled employment sectors, may be unduly discriminated against. The fact that native speakers of English may struggle to attain the highest levels in language tests such as IELTS, iBT TOEFL or on the CEFR has not been discussed sufficiently in research literature, nor the implications for migrants for whom English is an additional language. More research is needed in this area.

7.8.5 Recommendations regarding consequences for New Zealand

To ensure that test scores are used validly and ethically there is an urgent need for language testing researchers, test developers and policy makers to collaborate in order to clarify what is being tested and what can be inferred about test score results attained by native speakers of English as well as non-native speakers of English. Test developers have tried to do this by updating their websites (ielts, 2019) to include examples of the oral language that is likely to be produced at the most

common band scores (5.0, 6.0 and 7.0), but test users have to request a copy of the scores guide to read sample writing tasks at different band levels. My research shows that the test developers have failed to explain the meaning and connotations of test score figures for the majority of employers in New Zealand. For this reason, test developers have a responsibility to invest more time and money into disseminating more understandable information to test users (policy makers and employers). Further, it may be judicious and timely to devise a test of communicative competence for the Australasian context, rather than using a test, such as IELTS or iBT TOEFL, which measure higher language communication and cognition skills.

7.9 Conclusion

Collaboration, communication, creativity and critical thinking are widely reported as being the skills needed in today's workplace. However, there is very little collaboration and communication between test developers and test users, and the benefits of this may be undervalued by both parties. Spolsky (2008) and Kunnan (2018) suggested that the introduction of language testing into immigration procedures in many countries has resulted in the wilful misuse of language tests, particularly when governments ignore research. I suggest that it is also pertinent to question if test developers are aware of the differences between basic language cognition (BLC) and higher language cognition (HLC) and who is likely to be excluded by policies which use language test results because the current situation may be criticised as promoting unfairness.

Macqueen, Pill and Knoch (2016) called for test developers, providers and enforcers to ensure test effects are constructive. My research reinforces the need for more transparency regarding the type of written English language being assessed in large scale language tests compared with the written communication which is normally produced by native speakers of English. My data show that the majority of native speakers of English are not likely to produce highly complex, sophisticated, accurate written English and this type of English is not expected or needed in the contemporary workplace. Rather, native speakers of English are more likely to produce basic communicative written language which would score between 5.0 and 6.0 on a rating scale such as IELTS. Yet, immigration and professional registration bodies require non-native speakers of English to gain a score of 6.5 (overall score for skilled migrants) or 7.0 (in all four skills) on this rating scale. If native speakers of English would find it difficult to attain these scores in the reading and writing sections, as these research data show, immigration and professional registration policies are not being used in valid and ethical ways.

REFERENCES

AERA, APA, & NCME (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.

AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Retrieved from: http://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition

Al-Issa, A., Al-Bulushi, A. H., & Al-Zadjali, R. (2016). Arab English language teaching candidates climbing the IELTS mountain: A qualitatively driven hermeneutic phenomenology study. *The Qualitative Report*, *21*(5), 848-863.

Aryadoust, V. (2013). *Building a validity argument for a listening test of academic proficiency*. Newcastle upon Tyne, United Kingdom: Cambridge Scholars.

Bachman, L.F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly, 2*(1), 1-34.

Bachman, L.F., & Palmer, A.S. (2010). *Language assessment in practice*. Oxford, United Kingdom: Oxford University Press.

Barbour, R. (2007). *Doing focus groups*. London, United Kingdom: Sage Publications Ltd.

Basturkmen, H., & Elder, C. (2004). The practice of LSP. In A. Davies & C. Elder (Eds.), *Handbook of Applied Linguistics* (pp. 672-694). Oxford, United Kingdom: Oxford University Press.

Bogdan, R., & Biklen, S.K. (2007). *Qualitative research for education: An introduction to theories and methods*. Boston, MA: Pearson.

Bonk, W. J. (2018, July). *Effects of multiple repeated tests and skill trade-offs on an automatically scored speaking test*. Paper presented at the 40th Language Testing Research Colloquium, Auckland, New Zealand.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. doi: 10.1191/1478088706qp063oa

Brewer, C., Knoeppel, R.C., & Lindle, J.C. (2015). Consequential validity of accountability policy: Public understanding of assessments. *Educational Policy, 29*(5), 711-745. doi: 10.1177/0895904813518099

Brown, J.D. (2014). The future of world Englishes in language testing, *Language Assessment Quarterly, 11*(1), 5-26. doi:10.1080/15434303.2013.869817

Butorac, D. (2014). 'Like the fish not in water': How language and race mediate the social and economic inclusion of women migrants to Australia. *Australian Review of Applied Linguistics, 37*(3), 234-248.

Caines, J., Bridglall, B. L., & Chatterji, M. (2014). Understanding validity and fairness issues in high-stakes individual testing situations. *Quality Assurance in Education*, *22*(1), 5-18.

Canagarajah, S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly, 3,* 229-242.

Carroll, P.E., & Bailey, A.L. (2016). Do decision rules matter? A descriptive study of English language proficiency assessment qualifications for English language learners and native English speakers in fifth grade. *Language Testing, 33*(1), 23-52. doi:10.1177/0265532215576380

Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning, 10*, 807-815.

Cerbone, D. (2012). Phenomenological method: Reflection, introspection and skepticism. In D. Zahavi (Ed.), The Oxford handbook of contemporary phenomenology. Oxford, United Kingdom: Oxford University Press.

Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences. *Language Testing, 33*(4), 453-472. doi:10.1177/0265532215593312

Chalhoub-Deville, M., & Deville, C. (2005). A look back at and forward to what language testers measure. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 815-831). New York: Routledge. doi: 10.4324/9781410612700

Chalhoub-Deville, M., & Wigglesworth, G. (2005). Researcher report: Rater judgement and English language speaking proficiency. *World Englishes, 24*, 383-391.

Chapelle, C.A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics, 19,* 254-272.

Chapelle, C.A. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice, 29*(1), 3-13.

Chapelle, C.A. (2011). Validity argument for language assessment: The framework is simple…. *Language Testing, 29*(1), 19 –27. doi:10.1177/026553221141721

Chapelle, C.A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing, 32*(3), 385-405.

Chapelle, C.A., Enright, M.K., & Jamieson, J. (2008). Test score interpretation and use. In C.A. Chapelle, M.K. Enright, & J. Jamieson (Eds), *Building a validity argument for the test of English as a foreign language* (pp. 1-25). New York, NY: Routledge.

Chapelle, C.A., Enright, M.K., & Jamieson, J. (2010). Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice*, *29*(1), 3-13.

Cizek, G.J., Rosenbert, S.L., & Koons, H.H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68,* 397-412.

Coffey, A. & Atkinson, P. (1996). *Making sense of qualitative data*. London, United Kingdom: Sage Publications.

Cohen, L., Manion, L., & Morrison, K. (2009). *Research methods in education* (6th ed.). New York, NY: Routledge.

Cook, D.A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education, 49*, 560-575. doi: 10.1111/medu12678

Cronbach, L.J. (1988). Five perspectives on validity argument. In W. Howard & H. I. Braun (Eds.), *Test Validity* (pp. 1-14). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Cullen, J. (2005). The ethics of research in educational settings. In P. Adams., K. Vossler., & C. Scrivens. (Eds.), *Teachers' work in Aotearoa New Zealand* (pp. 252-261). Victoria, New Zealand: Thomson Dunmore Press.

Cummins, J. (1984) *Bilingual Education and Special Education: Issues in Assessment and Pedagogy.* San Diego, CA: College Hill.

Davidson, C. & Tolich, M. (1999). *Social science research in New Zealand: Many paths to understanding.* Auckland, New Zealand: Pearson Education.

Davies, A. (2011). Commentary: Does language testing need the native speaker? *Language Assessment Quarterly, 8*, 291-308.

Davies, A., & Elder, C. (2005). Validity and validation in language testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 795-813). New York, NY: Routledge. doi: 10.4324/9781410612700

Deygers, B. (2018, July). *Language and citizenship: A2 tests and low-educated learners.* Paper presented at the 40th Language Testing Research Colloquium, Auckland, New Zealand

Diem, S., & Young, M. (2015). Considering critical turns in research on educational leadership and policy. *International Journal of Educational Management, 29*(7), 838-850. doi: 10.1108/IJEM-05-2015-0060

Douglas, D. (2000). *Assessing languages for specific purposes: Theory and practice*. Cambridge, United Kingdom: Cambridge University Press.

Elder, C. (1993). How do subject specialists construe classroom language proficiency? *Language Testing, 10*(3), 235-254. doi:10.1177/026553229301000303

Elder, C. (2018, July). *Resisting changes to English testing requirements for Australian citizenship: Insights from a senate enquiry*. Paper presented at the 40th Language Testing Research Colloquium, Auckland, New Zealand.

Elder, C. & Brown, A. (1997). Performance testing for the professions: Language proficiency or strategic competence? *Melbourne Papers in Language Testing, 6*(1), 68-78.

Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics, 26*, 282-304. doi: 10.1017/S0267190506000146

Elder, C., & Kim, S.H.O. (2013). Assessing teachers' language proficiency. In A.J. Kunnan (Ed.), *The companion to language assessment* (pp. 454-470). Malden, MA: Wiley-Blackwell.

Elder, C., Harding, L., & Knoch, U. (2009). *OET reading revision study. Final report to the OET Centre. Melbourne*: University of Melbourne, Language Testing Research Centre.

Elder, C., & O'Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS English language requirements in skilled migration. *Alternative Law Journal, 36*(2), 110-115.

Ezzy, D. (2002). *Qualitative analysis: Practice and innovation*. St. Leonards, Australia: Allen & Unwin.

Gadamer, H. G. (2002). Truth and method (J. Weinsheimer & G. D. Marshall, Trans. 2nd revised ed.). New York, NY: Continuum.

Gadamer, H. G. (2014). Truth and Method (J. Weinsheimer & D. G. Marshall, Trans. revised 2nd ed.). London, United Kingdom: Bloomsbury.

Gibb, T. (2015). Literacy and language education: The quantification of learning. *New Directions for Adult and Continuing Education, 146*, 53 – 61. doi: 10.1002/ace.20131

Gill, P., Stewart, K., Treasure, E. & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal, 204*(6), 291-295.

Gordon, E.W. (2013). The Gordon commission's perspectives on the future of assessment. In Chatterji, M. (Ed.), *Validity and test use: An international dialogue on educational assessment, accountability and equity* (pp. 93-99). Bingley, United Kingdom: Emerald.

Gribble, C., Blackmore, J., Morrissey, A-M., & Capic, T. (2016). Investigating the use of IELTS in determining employment, migration and professional registration outcomes in healthcare and early childhood education in Australia. *IELTS Research Reports Online Series, 2016-2*. Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Hall, G. (2010). International English language testing: A critical response. *ELT Journal, 64*(3), 321-328. doi: 10.1093/elt/ccpo54

Hara, K. (1995). Quantitative and qualitative research approaches in education. *Education, 115*(3), 351-355.

Harrison, J., MacGibbon, L., & Morton, M. (2001). Regimes of trustworthiness in qualitative research: The rigors of reciprocity. *Qualitative Inquiry*, *7*(3), 323- 345. doi:10.1177/107780040100700305

Heidegger, M. (1967). *Being and time* (J. Macquarrie & E. Robinson, Trans.). Oxford, United Kingdom: Blackwell.

Hulstijn, J.H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Assessment Quarterly, 8*(3), 229-249. doi: 10.1080/15434303.2011.565844

Hulstijn, J. H. (2015). *Language proficiency in native and non-native speakers: Theory and research*. Amsterdam, Netherlands: John Benjamins.

IELTS 11 (2016). *IELTS Academic with answers. Authentic examination papers*. Cambridge, United Kingdom: Cambridge University Press.

Immigration New Zealand. (2019). *Acceptable English language test results: Skilled migrant category visa*. Retrieved from: https://www.immigration.govt.nz/new-zealand-visas/apply-for-a-visa/tools-and-information/english-language/acceptable-english-language-test-results/acceptable-english-language-test-results-smc

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3). doi:10.1177/0265532208090158

International English Language Testing System. (2019). *IELTS demographic data 2018*. Retrieved from: https://www.ielts.org/teaching-and-research/demographic-data

International English Language Testing System. (2019). *IELTS grows to 3.5 million a year*. Retrieved from: https://www.ielts.org/news/2019/ielts-grows-to-three-and-a-half-million-a-year

International English Language Testing System. (2019). *IELTS introduction*. https://www.ielts.org/what-is-ielts/ielts-introduction

International English Language Testing System. (2019). *Setting IELTS entry scores.* Retrieved from: https://www.ielts.org/ielts-for-organisations/setting-ielts-entry-scores

International English Language Testing System. (2019). Test taker performance 2018: Retrieved from: https://www.ielts.org/teaching-and-research/test-taker-performance

International English Language Testing System. (2019). *Who accepts IELTS scores? Find the score you need for your chosen organisation*. https://www.ielts.org/about-the-test/who-accepts-ielts-scores/ro/all/all/all/employer

International Language Testing Association (2018). *ILTA code of ethics*. Retrieved from: https://www.iltaonline.com/page/CodeofEthics

Janesick, V. (2003). The choreography of qualitative research design: Minuets, improvisations and crystallization. In N.K. Denzin & Y.S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 46-79). London, United Kingdom: Sage.

Jenkins, J. (2009). English as a lingua franca: Interpretations and attitudes. *World Englishes, (28)*2, 200-207. doi: 10.1111/j.1467-971X.2009.01582.x

Johnson, R.C. (2011). *Assessing the assessments: Using an argument-based validity framework to assess the validity and use of an English placement system in a foreign language context* (Doctoral thesis, Macquarie University, Australia). Retrieved from: http://hdl.handle.net/1959.14/284597

Johnson, R.C., & Riazi, A.M. (2017). Validation of a locally created and rated writing test used for placement in a higher education EFL program. *Assessing Writing, 32*, 85-104. doi: 10.1016/j.asw.2016.09.002

Jun, H.S. (2014). *A validity argument for the use of scores from a web-search-permitted and web-source-based integrated writing test* (Doctoral thesis, Iowa State University, USA). Retrieved from https://lib.dr.iastate.edu/etd/13899

Kachru, B.B. (1986). The power and politics of English. *World Englishes, 5*, 121-140.

Kane, M.T. (1992). An argument-based approach to validity. *Psychological Bulletin,112*(3), 527-535.

Kane, M.T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice, 21*(1), 31-41.

Kane, M.T. (2006). *Validation. Educational Measurement, 4*, 17-64.

Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*(1), 1-73.

Kane, M.T. (2016). Validity as the evaluation of the claims based on test scores. *Assessment in Education, 23*(2), 309-311.

Kim, H-J. (2006). World Englishes in language testing: A call for research. *English Today, 22*(4), 32-39.

Knoch, U., & Chapelle, C.A. (2018). Validation of rating processes within an argument-based framework. *Language Testing, 35*(4), 477-499. doi: 10.1177/0265532217710049

Knoch, U., & Macqueen, S. (2016). Language assessment for the workplace. In D. Tsagari, & J. Banarjee (Eds.), *Handbook of second language assessment* (pp. 291-307). Boston, MA: De Gruyter

Kunnan, A.J. (2005). Language assessment from a wider context. In E. Hinkel (Ed.), *Handbook of research in second language learning* (pp. 779-794). Mahwah, NJ: Lawrence Erlbaum.

Knoch, U., & Macqueen, S. (2020). *Assessing English for professional purposes*. Abingdon, United Kingdom: Routledge

Kunnan, A.J. (2018). *Evaluating language assessments.* New York, NY: Routledge

Lanteigne, B. (2006). Regionally specific tasks of non-western English language use. *TESL-EJ, 10*(2), 1-19.

Langridge, D. (2007) *Phenomenological Psychology*: Theory, Research and Method. Harlow, United Kingdom: Pearson

Lave, J., & & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, United Kingdom: Cambridge University Press.

Levinson, B., Sutton, M., & Winstead, T. (2009). Education policy as a practice of power. Theoretical tools, ethnographic methods, democratic options. *Educational Policy, 23*(6), 767-795. doi: 10.1177/0895904808320676

Li, Z. (2015). *An argument-based validation study of the English Placement Test (EPT) – Focusing on the inferences of extrapolation and ramification* (Doctoral thesis, Iowa State University, USA). Retrieved from https://lib.dr.iastate.edu/etd/14538

Lo Bianco, J. (2018, July). *No policy without testing! How the language of policy persuasion and persuasive language help to make testing count*. Paper presented at the 40th Language Testing Research Colloquium, Auckland, New Zealand.

Loewen, S. (2015). *Introduction to instructed second language acquisition*. New York, NY: Routledge

Lumley, T., & Brown, A. (2005). Research methods in language testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 833-855). New York, NY: Routledge. doi: 10.4324/9781410612700

Mackey, J. (2011). *New Zealand teachers' online professional development and communities of practice* (Unpublished doctoral thesis), Deakin University, Melbourne, Australia.

Macqueen, S., Pill, J., & Knoch, U. (2016). Language test as boundary object: Perspectives from test users in the healthcare domain. *Language Testing, 33*(2), 271-288. doi: 10.1177/0265532215607401

Macqueen, S., Yahalom, S., Kim, H., & Knoch, U. (2012). Exploring writing demands in healthcare settings. Unpublished report to the OET Centre. Melbourne: University of Melbourne.

Major, G., Terraschke, A., Major, E., & Setijadi, C. (2014). Working it out: Migrants' perspectives of social inclusion in the workplace. *Australian Review of Applied Linguistics*, *37*(3), 249-261.

McNamara, T. (1996). *Measuring second language performance*. London, United Kingdom: Longman

McNamara, T. (2000). *Language testing*. Oxford, United Kingdom: Oxford University Press.

McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick. *Language Assessment Quarterly, 3*(1), 21-21.

McNamara, T. (2008). The socio-political and power dimensions of tests. In N.H. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 2542-2554). Boston, MA: Springer. doi: 10.1007/978-0-387-30424-3_190

McNamara, T., & Shohamy, E. (2008). Language tests and human rights. *International Journal of Applied Linguistics, 18*(1), 89-95. doi: 10.1111/j.1473-4192.2008.00191.x

McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.

McNamara, T., & Ryan, K. (2011). Fairness vs justice in language testing: The place of English literacy in the Australian citizenship test. *Language Assessment Quarterly, 8*(2), 161-178.

Merrifield, G. (2008). An impact study into the use of IELTS as an entry criterion for professional associations – Australia, New Zealand and the U.S.A. *IELTS Research Reports, 8.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Merrifield, G. (2011). An impact study into the use of IELTS by professional associations and registration entities: Canada, the U.K. and Ireland. *IELTS Research Reports, 11.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Messick, S.J. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.

Messick, S.J. (1996). Validity and washback in language testing. *Language Testing,13(*3), 241-256.

Mishler, E. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review, 60*(4), 415-443.

Moore, T., Morton, J., & Price, S. (2011). Construct validity in the IELTS academic reading test: A comparison of reading requirements in IELTS test items and in university study. *IELTS Research Reports, 11.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Moore, T., Morton, J., Hall. D., & Wallis, C. (2015). Literacy practices in the professional workplace: Implications for the IELTS reading and writing tests. *IELTS Research Reports Online Series, 2015-1*. Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Murray, J., Cross, J.L., & Cruickshank, K. (2014). Stakeholder perceptions of IELTS as a gateway to the professional workplace: The case of employers of overseas trained teachers. *IELTS Research Reports Online Series, 2014-1.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

O'Loughlin, K. (2006). Assessment at the workplace. In E. Shohamy (Ed.), *Encyclopaedia of language and Education, Volume 7: Language Testing and Assessment*. New York, NY: Springer.

O'Loughlin, K. (2008). The use of IELTS for university selection in Australia: A case study. *IELTS Research Reports Online Series, 2008-8.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing, 30*(3), 363-142. doi: 10.1177/0265532213480336

O'Neill, R., Buckendahl, C.W., Plake, B.S., & Taylor, L. (2007). Recommending a nursing-specific passing standard for the IELTS examination. *Language Assessment Quarterly*, *4*(4), 295-317.

Ortega, L. (2013). *Understanding second language acquisition*. New York, NY: Routledge.

Patton, M.Q. (2001). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Pilcher, N., & Richards, K. (2017). Challenging the power invested in the International English Language Testing System (IELTS): Why determining 'English' preparedness needs to be undertaken within the subject context. *Power and Education, 9*(1), 3-17. doi: 10.1177/1757743817691995

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary enquiry. *Language Testing, 30*(3), 381-402. doi: 10.1177/0265532213480337

Polio, C., & Williams, J. (2009). Teaching and testing writing. In M. H. Long & C.J. Doughty (Eds.), *The handbook of language teaching* (pp. 486-517). Malden, MA: Wiley-Blackwell. doi:10.1002/9781444315783

Read, J., & Wette, R. (2009). Achieving English proficiency for professional registration: The experience of overseas-qualified health professionals in the New Zealand context. *IELTS Research Reports, 10.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Riazi, A.M. (2018, July). *Mixed methods research in language assessment validation studies*. Paper presented at the 40th Language Testing Research Colloquium, Auckland, New Zealand.

Ricento, T. (2014). Thinking about language: What political theorists need to know about language in the real world. *Language Policy, 13,* 351-369. doi: 10.1007/s10993-014-9322-2

Richardson, L. (1994). Writing: A method of inquiry. In N.K. Denzin and Y.S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 516-529). Thousand Oaks, CA: Sage.

Rose, H., & Syrbe, M. (2018). Assessment in teaching English as an international language. In J.I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1-12). Hoboken, NJ: Wiley. doi: 10.1002/9781118784235

Roshid, M.M., & Chowdhury, R. (2013). English language proficiency and employment: A case study of Bangladeshi graduates in Australian employment market. *Mevlana International Journal of Education, 3*(1), 68-81.

Rumsey, M., Thiessen, J., Buchan, J., & Daly, J. (2016). The consequences of English language testing for international health professionals and students: An Australian case study. *International Journal of Nursing Studies, 54*, 95-103.

Ryan, K. (2016). Language assessment and analysis for immigration, citizenship and asylum. In D. Tsagari, & J. Banarjee (Eds.), *Handbook of second language assessment* (pp. 309-323). Boston, MA: De Gruyter.

Sawyer, W., & Singh, M. (2012). Learning to play the 'classroom tennis' well: IELTS and international students in teacher education. *IELTS Research Reports Online Series, 11*(2), 1-54. Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Scarino, A. (2014). Situating the challenges in current languages education policy in Australia – unlearning monolingualism, *International Journal of Multilingualism, 11*(3), 289-306. doi: 10.1080/14790718.921176

Sedgewick, C., Garner, M., & Vicente-Macia, I. (2016). Investigating the language needs of international nurses: Insiders' perspectives. *IELTS Research Reports Online Series, 2016-2.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Sen, A. (2009). *The idea of justice*. Cambridge, MA: Harvard University Press.

Shohamy, E. (2001). *The power of tests: A critical perspective of the uses of language tests*. London, United Kingdom: Longman.

Shohamy, E. 2007. Language tests as language policy tools. *Assessment in Education*, *(14)*1, 117-130. doi:10.1080/09695940701272948

Shohamy, E., & McNamara, T. (2009). Language tests for citizenship, immigration, and asylum. *Language Assessment Quarterly, 6*(1), 1-5.

Smith, H., & Haslett, S. (2007). Attitudes of tertiary key decision-makers towards English language tests in Aotearoa New Zealand: Report on the results of a national provide survey. *IELTS Research Reports, 7.* Retrieved from: https://www.ielts.org/teaching-and-research/research-reports

Snape, D., & Spencer, L. (2003). The foundations of qualitative research. In J. Ritchie & J. Lewis (Eds.). *Qualitative research practice* (pp. 1-19). London, United Kingdom: Sage.

Spolsky, B. (2008). Language testing at 25: Maturity and responsibility? *Language Testing, 25*(3), 297-305. doi:10.1177/0265532208090153

Stacey, E. (2002). Social presence online: Networking learners at a distance. *Education and Information Technologies, 7*(4), 287-294.

Stevens, S. (2017). *The joy of movement in physical education: The enfleshed body* (Doctoral thesis, University of Canterbury, Christchurch, New Zealand). Retrieved from: https://ir.canterbury.ac.nz/handle/10092/14910

Stricker, L.J. (2004). The performance of native speakers of English and ESL speakers on the computer-based TOEFL and GRE general test. *Language Testing*, *21*(2), 146-173.

Talmy, S. (2011). The interview as collaborative achievement: Interaction, identity, and ideology in a speech event. *Applied Linguistics, 32*(1),25–42. doi:10.1093/applin/amq027.

Taylor, S. J., & Bogdan, R. C. (1998). *Introduction to qualitative research methods: A guidebook and resource* (3rd ed.). New York, NY: Wiley.

Taylor, S.J., Bogdan, R., & DeVault, M.L. (2016). *Introduction to qualitative research methods: A guidebook and resource* (4th ed.). Hoboken, NJ: Wiley.

Tolich, M & Davidson, C. (1999). *Starting fieldwork: An introduction to qualitative research in New Zealand*. Melbourne, Australia: Oxford University Press.

Toulmin, S. (1958). *The uses of argument*. Cambridge, United Kingdom: Cambridge University Press.

Toulmin, S. (2001). *Return to reason*. Cambridge, MA: Harvard University Press.

Uysal, H.H. (2010). A critical review of the IELTS writing test. *ELT Journal, 64*(3), 314-320. doi: 10.1093/elt/ccpo26

Van Moere, A. & Hanlon, S. (2018, July). *A Bayesian approach to improving measurement precision over multiple test occasions*. Paper presented at the 40th Language Testing Research Colloquium, Auckland, New Zealand.

Wengraf, T. (2001). *Qualitative research interviewing: Biographic, narrative and semi-structured methods.* London, United Kingdom: Sage.

Wette, R. (2011). English proficiency tests and communication skills training for overseas-qualified health professionals in Australia and New Zealand. *Language Assessment Quarterly, 8*, 200-210.

Yin, R. (2011). *Qualitative research from start to finish*. London, United Kingdom: The Guilford Press.

Appendix A: Ethical approval.

Appendix B: Ethical approval amendment.

Appendix C: Information letter for native speaker test takers.

Appendix D: Consent form for native speaker test takers.

Appendix E: Information letter for non-native speaker test takers.

Appendix F: Consent form for non-native speaker test takers.

Appendix G: Information letter for native speaker employers.

Appendix H: Consent form for native speaker employers.

Appendix I: Semi-structured questions for native speaker test takers.

Appendix J: Semi-structured questions for non-native speaker test takers.

Appendix K: Semi-structured questions for native speaker employers.

Appendix L: IELTS sample writing Task One.

Appendix M: IELTS sample writing Task Two.

Appendix N: IELTS sample speaking task for Part Two of the speaking test.

Appendix O: Task One writing descriptors (publicly available).

Appendix P: Task Two writing descriptors (publicly available).

Appendix Q: IELTS listening and reading mark conversions for the Academic and General Training modules.

Appendix R: Comparison of IELTS and CEFR band score results.

Appendix A: Ethical approval.

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2017/38

3 July 2017

Linda Edwards
School of Teacher Education
UNIVERSITY OF CANTERBURY

Dear Linda

The Human Ethics Committee advises that your research proposal "The validity and ethics of using IELTS for the purposes of immigration, enployment and professional registration in a NZ context" has been considered and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your emails of 2nd, 14th and 22nd June 2017.

Best wishes for your project.

Yours sincerely

pp. R. Robinson

Associate Professor Jane Maidment
Chair
University of Canterbury Human Ethics Committee

Appendix B: Ethical approval amendment.

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2017/38 Amendment 1

19 March 2018

Linda Edwards
School of Teacher Education
UNIVERSITY OF CANTERBURY

Dear Linda

Thank you for your request for an amendment to your research proposal "The validity and ethics of using IELTS for the purposes of immigration, enployment and professional registration in a NZ context" as outlined in your emails dated 21st February and 2nd and 13th March 2018.

I am pleased to advise that this request has been considered and approved by the Human Ethics Committee.

Yours sincerely

pp. R. Robinson

Professor Jane Maidment
Chair, Human Ethics Committee

Appendix C: Information letter for native speaker test takers.

Department: **Learning and Teaching Languages in the College of Education, Health and Human Sciences**
Telephone: +64 3 021 483 209

Email: linda.edwards@pg.canterbury.ac.nz

01.05.2017

Research title: The validity and ethics of using IELTS test scores for the purposes of immigration, employment and professional registration in a NZ context

Information Sheet for native-speakers of English taking the IELTS test

The main aim of this research is to make a detailed re-presentation of how decision makers or test-users perceive the International English Language Testing System (IELTS) and its consequences for migration in the hope that this will lead to a better understanding of language assessment and its relation to 'on the job' competence. You have been selected to be a participant because you are a native-speaker of English and are interested in taking the IELTS test because you are a migrant or out of educational/business interest.

My name is Linda Edwards and I am conducting doctoral research at the University of Canterbury into the use of IELTS for employment and migration in NZ. Since 1992, I have been an English language teacher, Director of Studies and the Director of Communicative Language Training International. In these roles, I have been involved with migrants who wish to live and work in NZ and take IELTS for the purpose of gaining the required band scores for employment/migration. I have also been an IELTS examiner since 1997. IELTS is a high stakes exam, which was originally developed for university entrance but is now being promoted as a testing mechanism for migration and employment in English speaking countries.

You will be asked to take the listening, reading and writing components of the test over the period of one morning. The speaking part of the exam will last for 15 minutes at a time that is convenient for you and may be conducted and recorded via video conferencing (Zoom). After taking the IELTS test, you will be sent a synopsis of the test results and asked to indicate if you can attend an interview session, lasting approximately 30 minutes in which you will be able to debate how you view the IELTS test, which parts you found easy or difficult, together with your perception of what is tested and the use of the exam for employment in an English-speaking country.

The interview discussions will be recorded using audio equipment. Any information you contribute in the interview sessions will be regarded as your personal opinions rather than views which are endorsed by any organizations you are associated with. You will be asked to

acknowledge that all information which you hear while involved in the research project is confidential and shall not be disclosed to any person except to those specifically authorised by the University of Canterbury or as is required by law.

In the performance of the tasks and application of the procedures there are risks that you may feel you have insufficient knowledge about the IELTS test. For this reason, you will be sent information about the test from the IELTS official website in advance. You may also feel that, as native-speakers of English, you should attain the highest IELTS band score (9.0). You will, therefore, be sent data from the IELTS website showing the results for candidates who indicate that their first language is English which will demonstrate that there are wide-ranging results for native-speakers of English. Other participants in the interview sessions will be native-speakers of English who have taken the test under the same conditions as you.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you, as long as this remains practically possible. However, once analysis of raw data starts on 10 January 2018 it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, the research data will be coded using pseudonyms, it will be securely stored on a server at the University of Canterbury. Rooms where the data are accessed will be locked while all computers and the files that are used for the research will be password protected. The people with access to the data will be the main researcher and her supervisors. If transcribers are used, they will sign the appropriate confidentiality forms. The data will be securely stored for 10 years and then destroyed. A thesis is a public document and will be available through the UC Library.

Please indicate to the researcher on the consent form if you would like to receive a copy of a summary of results of the project.

The project is being carried out as a requirement for a PhD by Linda Edwards under the supervision of Associate Professor Una Cunningham and Dr John Boereboom, who can be contacted at una.cunningham@canterbury.ac.nz and john.boereboom@canterbury.ac.nz. They will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return it to linda.edwards@pg.canterbury.ac.nz

Appendix D: Consent form for native speaker test takers.

Department
Telephone: +64 3 021 483 209
Email:
linda.edwards@pg.canterbury.ac.nz

Research project: The validity and ethics of using IELTS test scores for the purposes of immigration, employment and professional registration in a NZ context.

Consent Form for native-speakers of English taking the IELTS test

Please tick the appropriate boxes below:

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

☐ I understand that any information or opinions I provide will be kept confidential to the researcher and her supervisors and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

☐ I understand that the discussions in the interview sessions will remain private to protect the confidentiality of other participants. Any published or reported results from the interview sessions will not identify the participants.

☐ Any information contributed in the interview sessions will be regarded as personal opinion rather than views which are endorsed by any organizations I am associated with.

☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after ten years.

☐ I understand the risks associated with taking part and how they will be managed.

☐ I understand that I can contact the researcher, Linda Edwards at linda.edwards@pg.canterbury.ac.nz or supervisor, Associate Professor Una Cunningham at una.cunningham@canterbury.ac.nz for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)

☐ I would like a summary of the results of the project.

☐ By signing below, I agree to participate in this research project.

Name:_____ Signed:_____ Date:_____

Email address *(for report of findings)*:

_____

Please email this consent form to linda.edwards@pg.canterbury.ac.nz

Appendix E: Information letter for non-native speaker test takers.

Department: **Learning and Teaching Languages in the College of Education, Health and Human Sciences**
Telephone: +64 3 021 483 209

Email: linda.edwards@pg.canterbury.ac.nz

20.02.2018

Research title: The validity and ethics of using IELTS test scores for the purposes of immigration, employment and professional registration in a NZ context

Information Sheet for non-native-speakers of English taking the IELTS test

The main aim of this research is to make a detailed re-presentation of how decision makers or test-users perceive the International English Language Testing System (IELTS) and its consequences for migration in the hope that this will lead to a better understanding of language assessment and its relation to 'on the job' competence. You have been invited to discuss your perceptions of the IELTS test because you are a non-native-speaker of English and are interested in taking the IELTS test or you have already taken the test and now work in NZ.

My name is Linda Edwards and I am conducting doctoral research at the University of Canterbury into the use of IELTS for employment and migration in NZ. Since 1992, I have been an English language teacher, Director of Studies and the Director of Communicative Language Training International. In these roles, I have been involved with migrants who wish to live and work in NZ and take IELTS for the purpose of gaining the required band scores for employment/migration. I have also been an IELTS examiner since 1997. IELTS is a high stakes exam, which was originally developed for university entrance but is now being promoted as a testing mechanism for migration and employment in English speaking countries.

You will be asked to take the writing components of the test over the period of one hour if you have not previously done so. If you have taken the exam in the past you will be given exemplars of the test to refresh your memory of the test. This will help discuss how you view the IELTS test, which parts you found easy or difficult, together with your perception of what is tested and the use of the exam for employment in an English-speaking country.

The discussion will be in the form of a semi-structured interview and will be recorded using audio equipment. Any information you contribute will be regarded as your personal opinions rather than views which are endorsed by any organizations you are associated with. You will be asked to acknowledge that all information which you hear while involved in the research project is

confidential and shall not be disclosed to any person except to those specifically authorised by the University of Canterbury or as is required by law.

In the performance of the tasks and application of the procedures there are risks that you may feel you have insufficient knowledge about the IELTS test. For this reason, you will be sent information about the test from the IELTS official website in advance, which indicates that even native-speakers of English gain an average of 6.9 in the test. Individual test results will not be discussed.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you, as long as this remains practically possible. However, once analysis of raw data starts on 10 June 2018 it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, the research data will be coded using pseudonyms, it will be securely stored on a server at the University of Canterbury. Rooms where the data are accessed will be locked while all computers and the files that are used for the research will be password protected. The people with access to the data will be the main researcher and her supervisors. If transcribers are used, they will sign the appropriate confidentiality forms. The data will be securely stored for 10 years and then destroyed. A thesis is a public document and will be available through the UC Library.

Please indicate to the researcher on the consent form if you would like to receive a copy of a summary of results of the project.

The project is being carried out as a requirement for a PhD by Linda Edwards under the supervision of Associate Professor Alison Arrow, Dr. Jo Fletcher and Professor Letitia Fickel who can be contacted at alison.arrow@canterbury.ac.nz , jo.fletcher@canterbury.ac.nz and letitia.fickel@canterbury.ac.nz They will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return it to linda.edwards@pg.canterbury.ac.nz

Appendix F: Consent form for non-native speaker test takers.

Department
Telephone: +64 3 021 483 209
Email:
linda.edwards@pg.canterbury.ac.nz

Research project: The validity and ethics of using IELTS test scores for the purposes of immigration, employment and professional registration in a NZ context.

Consent Form for non-native speakers of English taking the IELTS test

Please tick the appropriate boxes below:

- ☐ I have been given a full explanation of this project and have had the opportunity to ask questions.
- ☐ I understand what is required of me if I agree to take part in the research.
- ☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.
- ☐ I understand that any information or opinions I provide will be kept confidential to the researcher and her supervisors and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.
- ☐ Any information contributed in the interview sessions will be regarded as personal opinion rather than views which are endorsed by any organizations I am associated with.
- ☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after ten years.
- ☐ I understand the risks associated with taking part and how they will be managed.
- ☐ I understand that I can contact the researcher, Linda Edwards at linda.edwards@pg.canterbury.ac.nz or supervisors, Associate Professor Alison Arrow at alison.arrow@canterbury.ac.nz, Associate Professor Jo Fletcher at jo.fletcher@canterbury.ac.nz or Professor Letitia Fickel at letitia.fickel@canterbury.ac.nz for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz)
- ☐ I would like a summary of the results of the project.

☐ By signing below, I agree to participate in this research project.

Name:_____ Signed:_____ Date:_____

Email address *(for report of findings)*:

_____

Please email this consent form to linda.edwards@pg.canterbury.ac.nz

Appendix G: Information letter for native speaker employers.

Department: **Learning and Teaching Languages in the College of Education, Health and Human Sciences**
Telephone: +64 3 021 483 209

Email: linda.edwards@pg.canterbury.ac.nz

01.05.2017

The validity and ethics of using IELTS test scores for the purposes of immigration, employment and professional registration in a NZ context

Information Sheet for professional decision makers

The main aim of this research is to make a detailed re-presentation of how decision makers or test-users perceive the International English Language Testing System (IELTS) and its consequences for migration in the hope that this will lead to a better understanding of language assessment and its relation to 'on the job' competence. You have been selected to be a participant because you are a professional who uses IELTS band scores for employment and/or migration decisions. For this reason, you are probably interested in gaining professional knowledge about what exactly is tested in the IELTS exam and how professionals in other fields make use of the exam results.

My name is Linda Edwards and I am conducting doctoral research at the University of Canterbury into the use of IELTS for employment and migration in NZ. Since 1992, I have been an English language teacher, Director of Studies and the Director of Communicative Language Training International. In these roles, I have been involved with migrants who wish to live and work in NZ and take IELTS for the purpose of gaining the required band scores for employment/migration. I have also been an IELTS examiner since 1997. IELTS is a high stakes exam, which was originally developed for university entrance but is now being promoted as a testing mechanism for migration and employment in English speaking countries.

You will be given an overview of the test and be given information on the results of native speakers of English who have taken the test for my research purposes. Secondly, you will be asked to give your opinions on language testing in a semi-structured interview lasting approximately 30 minutes in which you will be able to debate how you view the IELTS test, your perception of what is tested and the use of the exam for employment in an English-speaking country. All in all, the whole session will take 40-45 minutes.

Data will be collected from the session and analysed to see how people who use language testing as part of their work interpret and use test scores. For this reason, the discussions will be recorded using audio equipment. Any information you contribute in the interview will be regarded as your

personal opinion rather than views which are endorsed by any organizations you are associated with. You will be asked to acknowledge that all information which you hear while involved in the research project is confidential and shall not be disclosed to any person except to those specifically authorised by the University of Canterbury or as is required by law.

In the performance of the tasks and application of the procedures there are risks that you may feel you have insufficient knowledge about the IELTS test. For this reason, you will be sent information about the testing of native-speaker candidates in advance and the first 10 minutes of the session will focus on an overview of the contents of the test.

Participation is voluntary and you have the right to withdraw at any stage without penalty. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you, as long as this remains practically possible. However, once analysis of raw data starts on 10 January 2018 it will become increasingly difficult to remove the influence of your data on the results.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, the research data will be coded using pseudonyms, it will be securely stored on a server at the University of Canterbury. Rooms where the data are accessed will be locked while all computers and the files that are used for the research will be password protected. The people with access to the data will be the main researcher and her supervisors. If transcribers are used, they will sign the appropriate confidentiality forms. The data will be securely stored for 10 years and then destroyed. A thesis is a public document and will be available through the UC Library.

Please indicate to the researcher on the consent form if you would like to receive a copy of a summary of results of the project.

The project is being carried out as a requirement for a PhD by Linda Edwards under the supervision of Associate Professor Alison Arrow, Dr. Jo Fletcher and Professor Letitia Fickel who can be contacted at alison.arrow@canterbury.ac.nz , jo.fletcher@canterbury.ac.nz and letitia.fickel@canterbury.ac.nz . They will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the University of Canterbury Human Ethics Committee, and participants should address any complaints to The Chair, Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).

If you agree to participate in the study, you are asked to complete the consent form and return it to linda.edwards@pg.canterbury.ac.nz

Appendix H: Consent form for native speaker employers.

Department: **Learning and Teaching Languages in the College of Education, Health and Human Sciences**
Telephone: +64 3 021 483 209
Email:
linda.edwards@canterbury.pg.ac.nz,

Research project: The validity and ethics of using IELTS test scores for the purposes of immigration, employment and professional registration in a NZ context.

Consent Form for professional decision makers

Please tick the appropriate boxes below:

☐ I have been given a full explanation of this project and have had the opportunity to ask questions.

☐ I understand what is required of me if I agree to take part in the research.

☐ I understand that participation is voluntary, and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.

☐ I understand that any information or opinions I provide will be kept confidential to the researcher and her supervisors and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.

☐ I understand that the discussions in the interview sessions will remain private to protect the confidentiality of others participating in the focus groups. Any published or reported results from the focus groups will not identify the participants.

☐ Any information contributed in the interview sessions will be regarded as personal opinion rather than views which are endorsed by any organizations I am associated with.

☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after ten years.

☐ I understand the risks associated with taking part and how they will be managed.

☐ I understand that I can contact the researcher, Linda Edwards at linda.edwards@pg.canterbury.ac.nz or supervisors, Associate Professor Alison Arrow at alison.arrow@canterbury.ac.nz, Associate Professor Jo Fletcher at jo.fletcher@canterbury.ac.nz or Professor Letitia Fickel at letitia.fickel@canterbury.ac.nz for further information. If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-

ethics@canterbury.ac.nz)

☐ I would like a summary of the results of the project.

☐ By signing below, I agree to participate in this research project.

Name:_____ Signed:_____ Date:_____

Email address *(for report of findings)*:

_____

Please email this consent form to linda.edwards@pg.canterbury.ac.nz

Appendix I: Semi-structured questions for native speaker test takers.

1. What did you previously know about IELTS?

2. How did you find the test? Was the format and types of questions similar to what you expected?

3. What was easy for you? – with further probing into problems with macro and micro skills. Reading / writing/ speaking/ listening

Were the question types easy to follow?

Do you think these types of tasks are useful for your workplace/field of study?

Which tasks do you think show competence for study/work purposes?

4. What did you find particularly difficult? – again with further probing into problems with macro and micro skills.

Which question types were difficult to follow?

Which types of tasks did you feel were unusual?

How do you think you performed on the test?

5. How have your perceptions of the test changed or not changed?

6. What scores do you believe the majority of native speakers of English would get?

7. What advice do you have for people making decisions about band score requirements?

Appendix J: Semi-structured questions for non-native speaker test takers.

1. Tell me about your IELTS journey – your language(s) background, your test preparation, reasons for taking the test.

2. How did you find the test? What was easy/difficult for you? – with further probing into problems with macro and micro skills. Reading / writing/ speaking/ listening

3. What do you believe are your strengths/weaknesses with English. Were these reflected in the test results?

4. Do you think these types of tasks are useful and relate to competence for your workplace/field of study /immigration to NZ?

5. Which tasks or skills that are assessed are not relevant for your workplace/study purposes?

6. How would you compare your test scores with your workplace competence?

7. What are your perceptions of the links between test scores and workplace competence?

8. How have your test results affected your identity as a L2 speaker of English?

9. What scores do you believe the majority of native speakers of English would get?

10. What advice do you have for people making decisions about band score requirements?

Appendix K: Semi-structured questions for native speaker employers.

1. Tell me what you know about IELTS and its use for immigration purposes?

2. Do you use/not use IELTS in your work context?

3. Which language skills – reading, writing, speaking, listening – are most useful for on the job competence and social integration in your context?

4. What are the most common tasks that people need to perform in your context and what type of English do they need (formal/informal, simple/complex)?

5. What are the traits of someone who is competent in the workplace?

6. Now look at these charts – what do you believe a person who has a band score of 4.0, 5.0 or 6.0 be like in terms of English proficiency.

7. Which band scores do you think native speakers of English would receive?

   _____

8. Looking at these examples (writing tasks), which ones do you feel are competent users of English?

9. Listening to these people, are they competent speakers for your workplace?

10. Have a quick look at the reading and listening tasks and comment on what sort of language you believe is being tested.

11. Have a look at the results for native speakers of English and comment on what surprises you.

12. How have your perceptions of the test changed since reading the results of the native speaker group and their perceptions of IELTS assessment?

13. How confident do you now feel about your knowledge of the parameters of IELTS and how you (or other decision makers) are interpreting and using the band scores?

14. What advice do you have for people making decisions about band score requirements?

Appendix L: IELTS sample writing Task One.

Test 1

WRITING

WRITING TASK 1

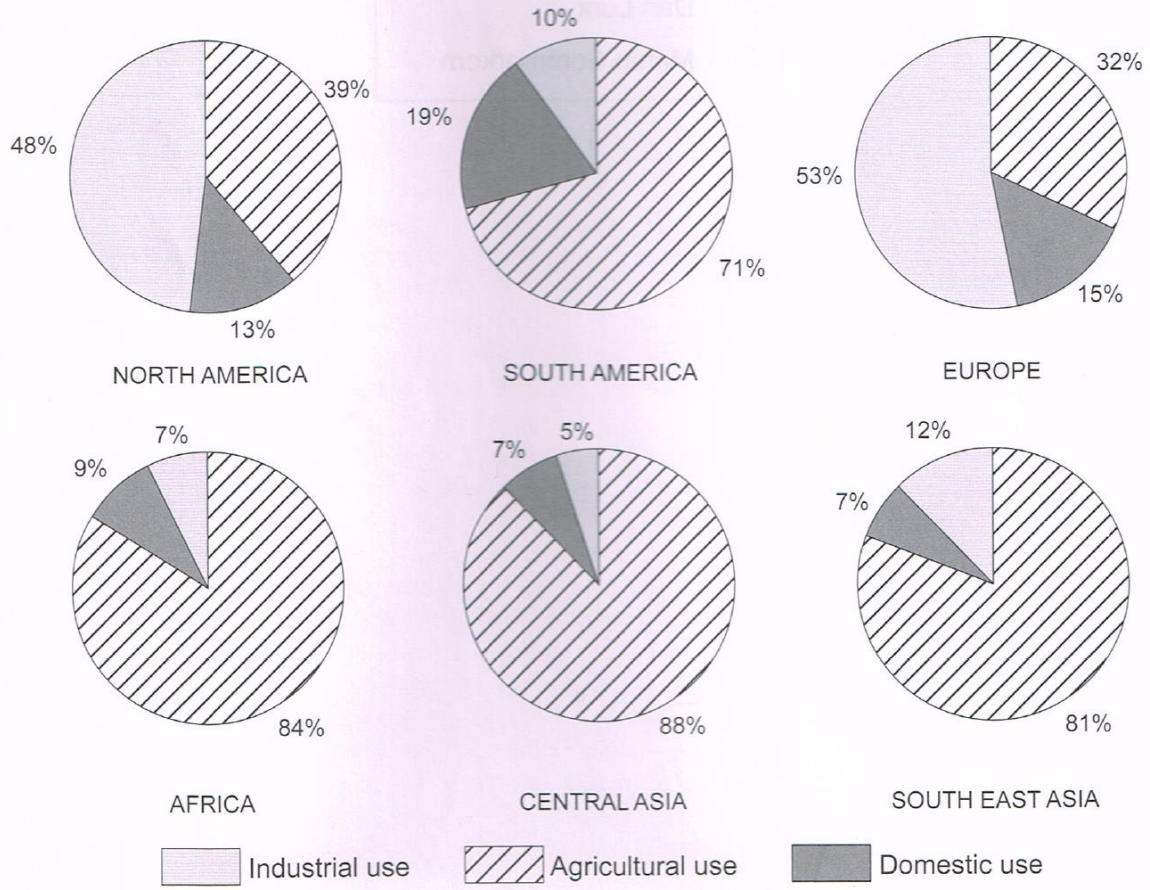
You should spend about 20 minutes on this task.

The charts below show the percentage of water used for different purposes in six areas of the world.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

Percentage of water used for different purposes in six areas of the world



30

Appendix M: IELTS sample writing Task Two.

Test 3

WRITING TASK 2

You should spend about 40 minutes on this task.

Write about the following topic:

> ***Some people say that the only reason for learning a foreign language is in order to travel to or work in a foreign country. Others say that these are not the only reasons why someone should learn a foreign language.***
>
> ***Discuss both these views and give your own opinion.***

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Appendix N: IELTS sample speaking tasks for the speaking test.

PART 1

The examiner asks the candidate about him/herself, his/her home, work or studies and other familiar topics.

EXAMPLE

Friends

- How often do you go out with friends? [Why/Why not?]
- Tell me about your best friend at school.
- How friendly are you with your neighbours? [Why/Why not?]
- Which is more important to you, friends or family? [Why?]

PART 2

Describe a writer you would like to meet.

You should say:
who the writer is
what you know about this writer already
what you would like to find out about him/her
and explain why you would like to meet this writer.

You will have to talk about the topic for one to two minutes. You have one minute to think about what you are going to say. You can make some notes to help you if you wish.

PART 3

Discussion topics:

Reading and children

Example questions:
What kinds of book are most popular with children in your country? Why do you think that is?
Why do you think some children do not read books very often?
How do you think children can be encouraged to read more?

Reading for different purposes

Example questions:
Are there any occasions when reading at speed is a useful skill to have? What are they?
Are there any jobs where people need to read a lot? What are they?
Do you think that reading novels is more interesting than reading factual books? Why is that?

56

Appendix O: IELTS Task One writing descriptors (publicly available).

IELTS™

WRITING TASK 1: Band Descriptors (public version)

Band	Task achievement	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	• fully satisfies all the requirements of the task • clearly presents a fully developed response	• uses cohesion in such a way that it attracts no attention • skilfully manages paragraphing	• uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	• uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	• covers all requirements of the task sufficiently • presents, highlights and illustrates key features/ bullet points clearly and appropriately	• sequences information and ideas logically • manages all aspects of cohesion well • uses paragraphing sufficiently and appropriately	• uses a wide range of vocabulary fluently and flexibly to convey precise meanings • skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation • produces rare errors in spelling and/or word formation	• uses a wide range of structures • the majority of sentences are error-free • makes only very occasional errors or inappropriacies
7	• covers the requirements of the task • (A) presents a clear overview of main trends, differences or stages • (GT) presents a clear purpose, with the tone consistent and appropriate • clearly presents and highlights key features/bullet points but could be more fully extended	• logically organises information and ideas; there is clear progression throughout • uses a range of cohesive devices appropriately although there may be some under-/over-use	• uses a sufficient range of vocabulary to allow some flexibility and precision • uses less common lexical items with some awareness of style and collocation • may produce occasional errors in word choice, spelling and/or word formation	• uses a variety of complex structures • produces frequent error-free sentences • has good control of grammar and punctuation but may make a few errors
6	• addresses the requirements of the task • (A) presents an overview with information appropriately selected • (GT) presents a purpose that is generally clear; there may be inconsistencies in tone • presents and adequately highlights key features/ bullet points but details may be irrelevant, inappropriate or inaccurate	• arranges information and ideas coherently and there is a clear overall progression • uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical • may not always use referencing clearly or appropriately	• •uses an adequate range of vocabulary for the task • attempts to use less common vocabulary but with some inaccuracy • makes some errors in spelling and/or word formation, but they do not impede communication	• uses a mix of simple and complex sentence forms • makes some errors in grammar and punctuation but they rarely reduce communication
5	• generally addresses the task; the format may be inappropriate in places • (A) recounts detail mechanically with no clear overview; there may be no data to support the description • (GT) may present a purpose for the letter that is unclear at times; the tone may be variable and sometimes inappropriate • presents, but inadequately covers, key features/ bullet points; there may be a tendency to focus on details	• presents information with some organisation but there may be a lack of overall progression • makes inadequate, inaccurate or over-use of cohesive devices • may be repetitive because of lack of referencing and substitution	• uses a limited range of vocabulary, but this is minimally adequate for the task • may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	• uses only a limited range of structures • attempts complex sentences but these tend to be less accurate than simple sentences • may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	• attempts to address the task but does not cover all key features/bullet points; the format may be inappropriate • (GT) fails to clearly explain the purpose of the letter; the tone may be inappropriate • may confuse key features/bullet points with detail; parts may be unclear, irrelevant, repetitive or inaccurate	• presents information and ideas but these are not arranged coherently and there is no clear progression in the response • uses some basic cohesive devices but these may be inaccurate or repetitive	• uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task • has limited control of word formation and/or spelling; • errors may cause strain for the reader	• uses only a very limited range of structures with only rare use of subordinate clauses • some structures are accurate but errors predominate, and punctuation is often faulty
3	• fails to address the task, which may have been completely misunderstood • presents limited ideas which may be largely irrelevant/repetitive	• does not organise ideas logically • may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	• uses only a very limited range of words and expressions with very limited control of word formation and/or spelling • errors may severely distort the message	• attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	• answer is barely related to the task	• has very little control of organisational features	• uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling	• cannot use sentence forms except in memorised phrases
1	• answer is completely unrelated to the task	• fails to communicate any message	• can only use a few isolated words	• cannot use sentence forms at all
0	• does not attend • does not attempt the task in any way • writes a totally memorised response			

Appendix P: IELTS Task Two writing descriptors (publicly available).

IELTS

WRITING TASK 2: Band Descriptors (public version)

Band	Task response	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	• fully addresses all parts of the task • presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas	• uses cohesion in such a way that it attracts no attention • skilfully manages paragraphing	• uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	• uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	• sufficiently addresses all parts of the task • presents a well-developed response to the question with relevant, extended and supported ideas	• sequences information and ideas logically • manages all aspects of cohesion well • uses paragraphing sufficiently and appropriately	• uses a wide range of vocabulary fluently and flexibly to convey precise meanings • skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation • produces rare errors in spelling and/or word formation	• uses a wide range of structures • the majority of sentences are error-free • makes only very occasional errors or inappropriacies
7	• addresses all parts of the task • presents a clear position throughout the response • presents, extends and supports main ideas, but there may be a tendency to over-generalise and/or supporting ideas may lack focus	• logically organises information and ideas; there is clear progression throughout • uses a range of cohesive devices appropriately although there may be some under-/over-use • presents a clear central topic within each paragraph	• uses a sufficient range of vocabulary to allow some flexibility and precision • uses less common lexical items with some awareness of style and collocation • may produce occasional errors in word choice, spelling and/or word formation	• uses a variety of complex structures • produces frequent error-free sentences • has good control of grammar and punctuation but may make a few errors
6	• addresses all parts of the task although some parts may be more fully covered than others • presents a relevant position although the conclusions may become unclear or repetitive • presents relevant main ideas but some may be inadequately developed/unclear	• arranges information and ideas coherently and there is a clear overall progression • uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical • may not always use referencing clearly or appropriately • uses paragraphing, but not always logically	• uses an adequate range of vocabulary for the task • attempts to use less common vocabulary but with some inaccuracy • makes some errors in spelling and/or word formation, but they do not impede communication	• uses a mix of simple and complex sentence forms • makes some errors in grammar and punctuation but they rarely reduce communication
5	• addresses the task only partially; the format may be inappropriate in places • expresses a position but the development is not always clear and there may be no conclusions drawn • presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail	• presents information with some organisation but there may be a lack of overall progression • makes inadequate, inaccurate or over-use of cohesive devices • may be repetitive because of lack of referencing and substitution • may not write in paragraphs, or paragraphing may be inadequate	• uses a limited range of vocabulary, but this is minimally adequate for the task • may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	• uses only a limited range of structures • attempts complex sentences but these tend to be less accurate than simple sentences • may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	• responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate • presents a position but this is unclear • presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported	• presents information and ideas but these are not arranged coherently and there is no clear progression in the response • uses some basic cohesive devices but these may be inaccurate or repetitive • may not write in paragraphs or their use may be confusing	• uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task • has limited control of word formation and/or spelling; errors may cause strain for the reader	• uses only a very limited range of structures with only rare use of subordinate clauses • some structures are accurate but errors predominate, and punctuation is often faulty
3	• does not adequately address any part of the task • does not express a clear position • presents few ideas, which are largely undeveloped or irrelevant	• does not organise ideas logically • may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	• uses only a very limited range of words and expressions with very limited control of word formation and/or spelling • errors may severely distort the message	• attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	• barely responds to the task • does not express a position • may attempt to present one or two ideas but there is no development	• has very little control of organisational features	• uses an extremely limited range of vocabulary, essentially no control of word formation and/or spelling	• cannot use sentence forms except in memorised phrases
1	• answer is completely unrelated to the task	• fails to communicate any message	• can only use a few isolated words	• cannot use sentence forms at all
0	• does not attend • does not attempt the task in any way • writes a totally memorised response			

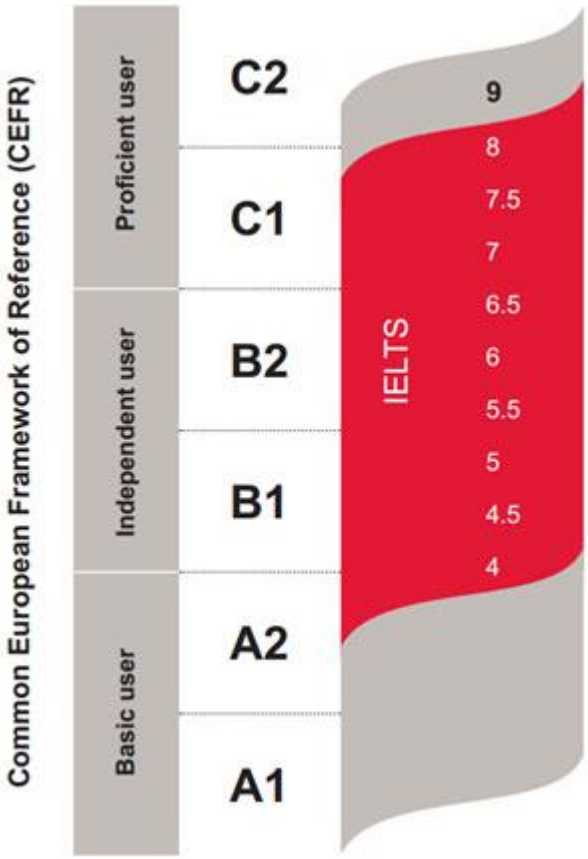
Appendix Q: IELTS listening and reading mark conversions for the Academic and General Training modules.

The tables below indicate the average number of marks required to achieve a particular band score in Listening, Academic Reading and General Training Reading.

Listening	
Band score	**Raw score out of 40**
5	16
6	23
7	30
8	35

Academic Reading		General Training Reading	
Band score	**Raw score out of 40**	**Band score**	**Raw score out of 40**
5	15	4	15
6	23	5	23
7	30	6	30
8	35	7	34

Appendix R: Comparison of IELTS and CEFR band score results.



Questions about IELTS and the CEFR

1. Has the IELTS test changed?

No, the test has not been changed.

2. Why is IELTS changing the way the band scores relate to the CEFR?

We have always been committed to providing ongoing revision as we grow in our understanding of the relationship between IELTS, other examinations and the CEFR levels.

The CEFR is becoming more prominent in how institutions consider language ability requirements. It is important therefore that we provide updated advice as to how to interpret IELTS scores in CEFR terms. The table previously on the website did not show half-band scores, and predated the introduction of half-band reporting for Writing and Speaking in July 2007.

3. Has IELTS been made more difficult?

No the test has not changed. The way it is examined and the way band scores are awarded

remain the same.

4. Should institutions and organisations which use IELTS scores change the band scores they expect students to achieve as a result of the revised CEFR mapping?

The test has not changed and the performance represented by each band score remains the same. The IELTS Scores Explained DVD provides samples of those performances so that institutions can judge what level is appropriate to their needs. There is no need for institutions to make changes where they have previously been satisfied with their particular score requirements.

5. Some IELTS band scores are shown as borderline (e.g. it is not clear whether band 5 is B1 or B2). How should institutions and organisations interpret this?

As IELTS preceded the CEFR, IELTS band scores have never aligned exactly with the CEFR transition points. The new table makes this clearer. Previously we provided advice as to the score on IELTS that a test taker who was at a given CEFR level might achieve. However, our research shows that a C1 minimum threshold would fall between the 6.5 and 7 bands on the IELTS scale. Therefore, whilst many 6.5 test takers would be at C1, a number will be marginally below. So if an institution requires a high degree of confidence that an applicant is at C1, they may wish to set a requirement of 7, rather than 6.5.

6. Does IELTS differentiate at C2 level?

Band scores of 8.5 and higher are recognised as C2. Band 8 is borderline.

7. If a student already has an IELTS score of 6.5, shown as C1 in the previous mapping, should this now be treated as a B2 equivalent score?

The score 6.5 is borderline B2/C1. The real-world level of performance represented by the result has not changed. It is for institutions to decide whether they wish to change their requirements, if alignment to a particular level of the CEFR is critical (see response to q5 above). The advice in the IELTS Guide for Educational Institutions as to probable levels required for different types of course still holds.

https://www.ielts.org/ielts-for-organisations/common-european-framework